2019

Snap Scholar: The User Experience of Engaging with Academic Research Through a Tappable Stories Medium

Ieva Burk

Claremont McKenna College

Snap Scholar: The User Experience of Engaging with Academic Research Through a Tappable Stories Medium

submitted to
Professor Mark Huber
and
Professor Alexandra Papoutsaki

by
Ieva Burk

for
Senior Thesis
Spring 2019
April 29, 2019

Acknowledgements

This thesis would not have been possible without the inspiration and support from an indispensable group of individuals. Thank you, to:

- the women in technology who inspire me most—Pamela Palmer and Farnaz Ronaghi—for providing me the space and experience to have the fulfilling summer that inspired this project, and for helping me see what I am capable of;
- my mentors at NovoEd for motivating me to make strides in education technology and for giving me the opportunity to work on their product this summer;
- the pioneers of Stanford Scholar [50]—Rajan Vaish, Amin Saberi, Shirish Goyal, and Sharad Goel—for paving the way for this research and for their enthusiasm in allowing me to join the initiative;
- my guide for this project, Rajan Vaish, for letting me come visit the office at Snap Inc. to get feedback on my designs, and for providing me the resources to acquire the skills I needed for making this work happen;
- my teachers at the Hive—Kareem Collie, Shannon Randolph, and Fred Leichter—for showing me the thoughtful approach that meaningful design requires, and for instilling in me a deep passion for human-centered design;
- my reader and professor Mark Huber, for his enthusiasm for my proposed topic, the flexibility of his support through both the technical and soft parts of my project, and for his general unwaning kindness;
- my reader Alexandra Papoutsaki, for nurturing my blooming interest in HCI, and supporting me through minor setbacks and moments of uncertainty;
- my apartment-mates Namrata Dev and Lachlan Montgomery, for listening to my late-night design rants for this project, and encouraging me to live and breathe the user experience;
- my friends Paloma and Omar, for endlessly supporting my endeavors, and building upon my crazy ideas with me;
- my friends Jason Barbarosh and Edgar Warnholtz, for their company and support during grinding periods, and always having espresso and Oatly on deck for Poppa breaks;
- my incredible mother, whose everyday sacrifices are the reason I am able to get an exceptional education;
- and the participants of my study, for leaving thoughtful and candid responses, and making this research possible.

This project is a culmination of my strife for understanding the end user and helping others uncover new passions.

Contents

1 Introduction

With the shift to learn and consume information through our mobile devices [44], most academic research is still only presented in long-form text. The Stanford Scholar Initiative [20, 50] has explored the segment of content creation and consumption of academic research through video. However, there has been another popular shift in presenting information from various social media platforms and media outlets in the past few years. Snapchat and Instagram have introduced the concept of tappable “Stories” that have gained popularity in the realm of content consumption, as shown in Figure 1 and Figure 2. Figure 3 breaks down the proportions of how much these services have gained popularity in the realm of content consumption among teenagers and young adults more recently.

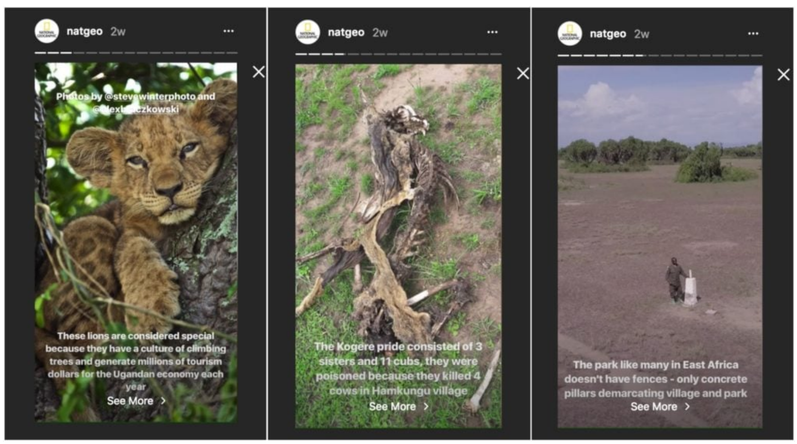


Figure 1: NatGeo posting Stories content on their Instagram [4].

To accelerate the growth of the creation of these research talks, I propose an alternative to video: a tappable Snapchat-like interface. This style is achieved using AMP [5], Google’s open source project to optimize web experiences on mobile, and particularly the AMP Stories visual medium [6]. My research seeks to explore how the process and quality of consuming the content of academic papers would change if instead of watching videos, users would consume content through Stories on mobile instead.

Since this form of content consumption is still largely unresearched in the academic context, I approached this research with a human-centered design process, going through a few iterations to test various prototypes before formulating research questions and designing an experiment. I tested various formats of research consumption through Stories with pilot users, and learned many lessons to iterate from along the way. I created a way to consume research papers in a Stories format, and designed a comparative study to measure the effectiveness of consuming research papers through the Stories medium and the video medium.

The results indicate that Stories are a quicker way to consume the same content, and improve the user’s pace of comprehension. Further, the Stories medium provides

Figure 2: An example of a Snapchat Story with filters and stickers.

the user a self-paced method—both temporally and content-wise—to consume technical research topics, and is deemed as a less boring method to do so in comparison to video. While Stories gave the learner a chance to actively participate in consumption by tapping, the video experience is enjoyed because of its reduced effort and addition of an audio component. These findings suggest that the Stories medium may be a promising interface in educational contexts, for distributing scientific content and assisting with active learning.

2 Related Work

In this section, I present an overview of prior work done in education technology and mediums of content consumption. I highlight the prior research related to mobile consumption, the use of video in the educational context, as well as the introduction of the tappable "Stories" medium by various social networks, which was a primary motivator for this work. My research seeks to explore what the Stories platform may mean in the context of educational content consumption.

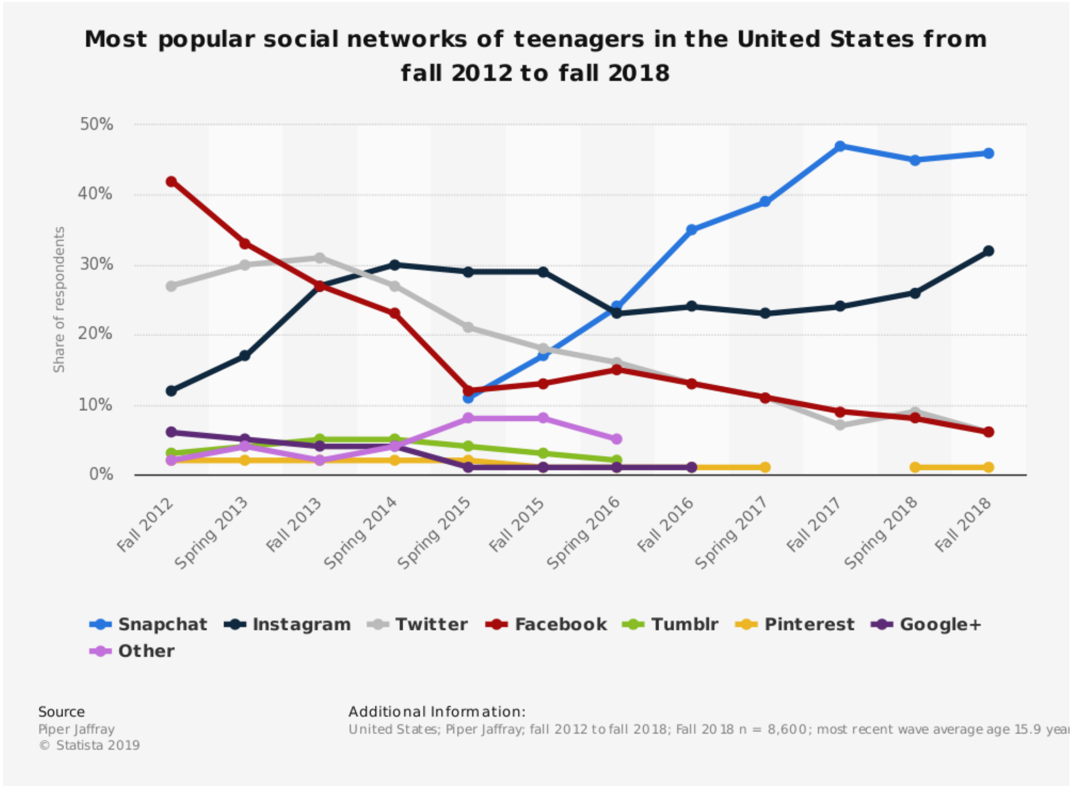


Figure 3: Preferred social networks of U.S. teens from 2012 to 2018 (Fall 2012 to Fall 2018; n = 8600; most recent wave average age of 15.9 years) [46].

2.1 Video

There is an increasing demand to consume information through video [44, 34]. Research based on responses from a 30-minute nationally representative online survey of 2587 respondents, 14 to 40 years old in 2018 found that video takes a significant role in the youngest generation's educational experience, with 59% of respondents ages 14-23 stating that YouTube is their preferred learning method, closely followed by Instagram and Snapchat [41]. The majority of academic research, however, is still only presented in text format, likely because of the limited time, resources, and lack of incentives [50] for researchers to convert their research content into other mediums. This gap between the way we consume information and the way academic literature is presented narrows the accessibility of research content in terms of its approachability and reach, and may discourage younger audiences from consuming research altogether.

The shift to learn and consume information through multimedia has partly been addressed by massive open online courses (MOOCs), which provide learning materials in a video-lecture format. Other efforts such as YouTube channels "Khan Academy" [29] and "Crash Course" [22] are known for turning tough topics into fast-paced, engaging videos. In the realm of academic research specifically, channels such as "Two Minute Papers" [52] and "Papers We Love" [2] have provided a quick way for viewers with limited technical expertise to gain a basic understanding of various computer science topics in video format.

2.2 Stanford Scholar

Still, the vast majority of academic research is only distributed in long-form text. To address this disparity, a team of Stanford researchers [50] introduced and evaluated an end-to-end system for creating 5-minute research talks developed collaboratively by volunteers worldwide.

The research videos were initially produced by collaborative teams of individuals, and could be subsequently edited or improved by any keen participants. The collective endeavor became known as the Stanford Scholar Initiative [20], an effort to make research more accessible and engaging while simultaneously providing a collaborative platform for teams of volunteers who are interested in learning more about a particular research topic by creating short talks about a paper in its field. The initiative further increased the accessibility of technical knowledge by converting English-language papers into video talks in multiple languages, and was structured in a way that did not involve the paper authors or subject matter experts [50].

There were a few challenges in crowdsourcing volunteers to collaborate on a creative endeavor. The first was to find a way to collaboratively produce videos that can be edited into a complete research presentation. To address this, Vaish et al. introduced a new approach to collaborative video creation. They drew on research in micro-tasking [8] and standardized each talk to consist of three components: slides, a written script, and voice-overs. Those components were then stitched together through a program that generated a complete video. In order to streamline the process, they built an online tool that allows crowd volunteers to collaborate and record audio slide-by-slide. Their modular approach supported quick editing and decreased retake time, both during and after the initial videos were created.

The next challenge involved facilitating the complex coordination between the large groups of crowd volunteers of inevitably varying expertise. To address this, they drew on previous research in effective team coordination by designing a structured scaffolding process to coordinate volunteers. Specifically, they divided the talk creation process into three discrete phases spanning a period of 21 days (three weeks) as shown in Figure 4: (1) on-boarding the crowd and forming teams; (2) generating a slide deck that includes both the talk slides and a slide-by-slide script of the talk; and (3) converting the script to slide-by-slide audio recordings, and reviewing the complete video presentation.

They issued an open call for participation in the summer of 2016 to evaluate their proposed system and ended up gauging interest from 840 people in 52 different countries. The crowd volunteers produced 219 videos completely on their own. To evaluate the talks on the consumption end, they gathered responses from 73 authors of converted papers and 300 outside reviewers. The video talks were frequently rated as "very good" and "very useful," receiving a median score of 4 out of 5 on both metrics of quality and utility. The results suggested that their approach was a scalable and

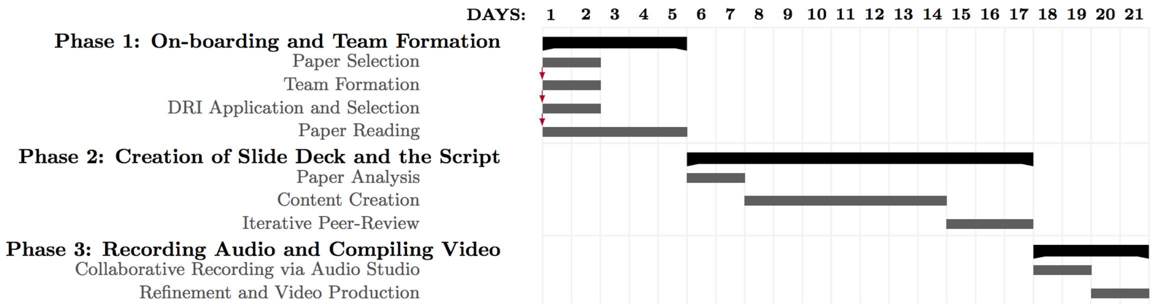


Figure 4: The three-phase structured workflow for crowdsourced video talk creation on the Stanford Scholar platform [50].

promising way to produce high-quality research content and increase the distribution and accessibility of scientific ideas.

With video's gain in popularity, it is important to study user behavior around video consumption. Various works have shown that video consumption is nonlinear, and have introduced the concept of video skimming [49, 12, 11] which suggests that the way we navigate videos may be similar to the ways we consume information in text format. The sporadic consumption that occurs when seeking information is especially important to consider for content creation and interface design.

2.3 Mobile Consumption

The prevalence and convenience of mobile has caused a shift in consumer behavior that presents new challenges and opportunities for usability. Social media platforms like Snapchat, Instagram, and Facebook have evolved the way people communicate and consume information through mobile. The time people are spending on mobile is increasing: their mobile device is most likely to be the first screen they interact with when they wake up and the last one they check before going to bed [17].

Users are not only consuming more content on mobile, but are also processing it faster than on a desktop computer. Facebook's research team recruited users who did not have a creative or advertising background to watch and evaluate their experiences watching video ads in News Feed. Their findings suggested that users are able to consume information faster on mobile in 1.7 seconds versus 2.5 seconds on a desktop, with younger audiences scrolling through more quickly than the average. Further, they are able to recall the information in 0.25 seconds on mobile [16].

The content that users are consuming may be the same across devices, but the user's varying behavior across mediums impacts how they process the content. Mobile consumption diverges from the traditional linear content experience since users are able to move back and forth through content, and are not staying in any page for any significant amount of time.

The proliferation of cameras on mobile devices has evolved the communication style of mobile to become more and more visual. With the rapid adoption of social networks, photo and video have become a ubiquitous part of the mobile experience. Currently, video consumption among social networks is primarily sound-off, "meaning that a video needs to visually attract an audience and, in most cases, communicate without sound" [17]. In a random sample of 800 ads from large Facebook advertisers in 2016, findings indicated that videos that required sound to understand the messaged dropped from 35% to 29% from the beginning of Q1 in January 2016 to the end of Q2 in June 2016 [16]. This raises concerns in accessibility, particularly how soundless video can be experienced by visually-impaired users.

2.4 Snapchat

Since its release in September 2011, Snapchat has grown to become one of the most popular social media apps with over 180 million daily users [3]. Originally known for introducing the concept of ephemeral content-sharing, Snapchat has extended its applications to include several new features and methods of content-sharing. It has primarily distinguished itself in the social network ecosystem through representing a new, mobile-first direction for social media.

In October 2013, Snapchat introduced the Stories feature, as demonstrated in Figure 2, which expanded on its original focus of one-to-one photo and video sharing by allowing users to post series of photos and videos that can then be viewed by friends for a 24-hour period [15]. Eight months later, the Stories feature had surpassed one-to-one snaps as the most frequently-used function of Snapchat, with over one billion Stories viewed per day [23].

In January 2015, Snapchat released the Discover tab, a channel-like feature within the app which allowed major brands and publishers to produce ad-supported short-form content in the Stories format. Using the unique tappable presentation style of its Stories feature, Discover allowed Snapchat to offer its own distinct publication, media, and news content in-app. In May 2017, it was reported that Snap Inc., Snapchat's parent company, signed deals with the NFL, Vice Media, Discovery, NBCUniversal, ABC, and BBC among others to produce original content for Snapchat in the Stories format [43].

2.5 Instagram

In August 2016, Instagram launched Instagram Stories, highly reminiscent of the Stories feature on Snapchat, as shown in Figure 1. The feature cloned the fleeting nature of Snapchat Stories that expire after 24 hours, and was incorporated into Instagram's feed to show Stories of people the user is following. Stories seemed to integrate seamlessly into Instagram's existing one-to-many model of sharing, but at the cost of less control over who views your Stories. The intimacy that was once unique to Snapchat changed in November 2018 when Instagram released Close Friends. This feature allowed users to customize their Stories audience, building onto Instagram Stories' existing functionality of content-sharing by adding flexibility to share moments with a smaller group of the user's choice.

2.6 The AMP Project

AMP [5] is a Google-run open source initiative to improve the performance of the web content and ads through optimizing web pages for delivery on mobile devices. AMP originated from Google in February 2016 when it became clear that the experience of browsing through certain mobile web pages was slow and clunky. The problem was not necessarily that the technology was lacking– there was a way to build experiences on mobile, but it required the right knowledge, resources, and web support. Rather, it was a lack of a framework to make a seamless web-based experience on mobile. Thus, the goal of AMP became reenvisioning the web to be well-documented, easily-deployable, validatable, and knowledgeable about user-first principles.

2.6.1 AMP Stories

More recently, users have started to consume more and more content on their mobile devices than on websites [36]. Although users read long-form articles on mobile, they only read them for an average of 1-2 minutes. The long-form article format makes content consumption more difficult for users since there is no clear way to find which sections they may want to do a deeper dive on. After testing with a small group of publishers for a few months, Google released AMP Stories in February 2018 to provide publishers with new storytelling options for the mobile web [18]. AMP Stories are just AMP pages stitched together into a tappable visual narrative with images, videos, graphics, and audio, as shown in Figure 5. AMP Stories allows users to create more permanent content in the story format that is not limited by the ephemeral timeline of its social media counterparts. While AMP Stories was originally optimized for mobile, it was designed to be adaptive to other user interfaces, as demonstrated in Figure 6. Publishers can create a story for mobile and have this automatically be adapted to work on landscape displays as well [35].

A Story is made up of one or more pages in which the user can tap through the story's pages by clicking on the right side of the screen, and go back to the previous page by tapping on the left side of the screen. The interaction is similar to navigating through Snapchat or Instagram stories.

Each page of a story is represented by an AMP story page tag. AMP story pages are made up of multiple layers stacked on top of one another in order to create a desired visual design. At the end of the story the system automatically appends a templated UI called a bookend in which the user can continue an onward journey.

3 Evaluation

Usability is most often defined as the potential to accomplish the goals of the user or the ease of use of an interface. In user-centered design, the evaluation of interfaces typically occurs before higher fidelity prototypes of a software project are developed. When usability evaluation is carried out at the end of the design cycle, changes become costly and harder to implement. Both usability inspection methods and usability

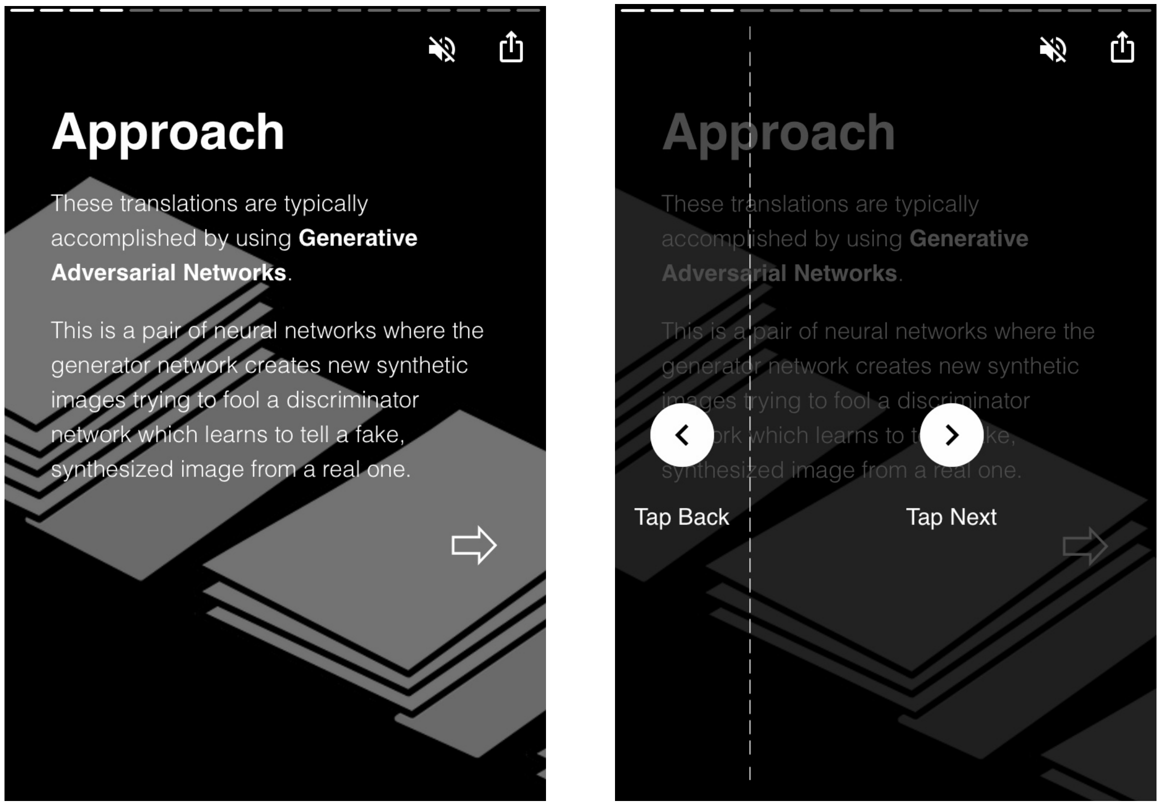


Figure 5: The AMP Story interface on mobile.

Figure 6: AMP Stories can also be adapted to desktop displays.

test methods are used to evaluate interfaces. Usability inspection methods are more informal and focus on cost-effective ways to find usability problems whereas usability

Figure 7: AMP Stories are composed of individual pages. These pages are further composed of individual layers that contain basic HTML and AMP elements [42].

Figure 8: Sample slides [50] from research talks produced in Japanese, English, Oriya, Chinese, Spanish and Hindi (from left to right, top to bottom). 67 of the 107 research talks produced were in foreign languages.

testing is more focused on testing with end users to provide direct information about the exact problems people face when using a system [25].

3.1 Usability Inspection Methods

Heuristic evaluation is the most widely used informal usability evaluation method. In a heuristic evaluation, a small set of evaluators examine a user interface and look for violations of a pre-determined set of design principles [38]. Evaluators will originally inspect the interface alone, and only after all evaluators have completed their inspection are they allowed to communicate and accumulate their findings. The restriction is meant to reduce bias and ensure independent evaluations. During an evaluation session, the evaluator will use the interface several times, examine the interactive elements, and compare them with a list of established usability principles, such as Nielsen's 10 usability heuristics [38]:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Recognize, diagnose, and recover from errors
10. Help and documentation

Advantages of heuristic evaluation are that it is quick, intuitive, useful throughout the development process, and employs practical identification of both minor and major problems. Disadvantages are that it does not necessarily fit the evaluation for all interface formats (desktop, mobile, etc.), the rigid nature of predetermined heuristics, difficulty in identifying or allowing for unknown user needs, and that evaluators can focus too much on one section.

Cognitive walkthrough is another discount usability method which is used to evaluate interfaces without end users. It is commonly employed when a full prototype is not feasible, and is based on formal mental models that assume that users learn to use an interface through complex guessing strategies [14]. The method involves walking through a user scenario step-by-step, dissecting the user's goal, and thinking through why they would take a certain next action.

Advantages of cognitive walkthrough are making end goals and expectations of the user explicit, independence from end users, use of low-fidelity or concept prototypes, and the analysis of potential user behavior can build empathy of the user's perspective. Disadvantages include no involvement with end users, heavy emphasis on the details of a particular task rather than an entire user interface, and potential bias in choosing the tasks.

Action analysis is a more formal method focused on what the user does rather than what they say they do. Also known as keystroke-level analysis, action analysis involves dividing a user task up into individual components such as typing on the keyboard or clicking on a video, and calculating the times taken to perform the associated actions. Altogether these actions comprise to achieve a desired goal on the interface [25].

Advantages are that it gives a precise metric for time it takes user to complete a task, its objectives, and breaking down an end goal into tasks encourages smarter design. Disadvantages of the method are that it is time-consuming, and there is no insight into whether or not there is a better set of tasks to achieve the user's goal.

3.2 Usability Test Methods

Thinking aloud is a test method that involves having an end user continuously verbalize their reactions to an interface. The method is useful since it gives insight into how the system is perceived by the user in real time, and surfaces any major misconceptions about how to use various aspects of the interface. When the user gives feedback in real time, they are verbalizing the contents of their working memory, which is much more accurate than retroactive reporting [10]. An alternate form of the thinking aloud technique is co-discovery in which two participants work together to perform a task and are encouraged to talk to each other as they work [27]. Talking through an interface with someone else may be more natural than thinking out loud, and so co-discovery often yields more and richer insights into what users are thinking when completing a task. Although the collaboration allows participants to process the interface in a more natural way, the disadvantages to co-discovery are that there may be bias in perception: one participant who quickly catches on to the interface could influence how the other participant interprets the features.

Advantages of thinking aloud is that it reveals user intent and thought process, provides a large collection of data from a relatively small user base, produces vivid user reactions, and helps surface misconceptions to anticipate future design problems. Disadvantages are that it is time-consuming, expensive, and can be perceived as unnatural and strenuous when performed alone since the observer asks the user to focus on a task until they complete it.

Field observation is a test method takes place in the user's context rather than in an observer's office or lab setting. Field observation can allow the observer to understand how the interface is used in a particular setting, and do so in a non-obstructive way. Therefore, it is typically performed with minimal intervention in order to produce the most genuine results. The observer should ideally be invisible to the participant to further ensure comfortable and natural conditions. In extreme cases, video recording can be asked to be used in order to deeply analyze the user's experience while being as discrete as possible [25].

Advantages of field observation are that it reveals natural use of the interface, provides authentic insights from user's real life setting, and encourages contextual design. Disadvantages are that it requires more than 20 users, can be expensive to travel to the user, and is only applicable for late stage prototypes.

Surveys and questionnaires are often used to translate subjective user feedback into more measurable insights. Surveys allow for quick feedback from a large response pool that can be statistically analyzed without much extra effort. These methods are indirect, however, since the users are not evaluating the actual user interface; they are only expressing their opinions about the interface. It is generally agreed upon that "data about people's actual behavior should have precedence over people's claims of what they think they do" [25].

Advantages of surveys and questionnaires are that they are quick, inexpensive, can be conducted at scale, make it easier to get a diverse user set, allow for consistent questions among participants, and deliver responses that can be statistically compiled.

Disadvantages include the disparity between user claims and user actions, collections of data about subjective measures, and potential of biased questions.

A/B testing is an interface optimization method that involves randomly splitting the traffic to your interface between two or more versions. User engagement metrics are then gathered and used to form the decision of whether or not the feature will be adopted into the interface. The method can be employed when there is a dilemma in choosing between two conflicting elements of an interface. A/B or split testing is beneficial in evaluating proposed solutions for problems that have already been identified in the interface [31].

Theoretically, we would want to run an experiment with an equal amount of users across the conditions. However, in the context of applications in industry, the biggest concern in A/B testing is testing a potentially disastrous idea with users that would dissuade them from using the product altogether. The best practice within industry is to expose a few users to the new idea, and choose the next course of action accordingly based on the continuous stream of feedback. If there is no major negative effect, we can dial up the amount of users to test on and eventually expose 50% of users to each test. If there is a major negative effect, this allows us to abort the experiment and analyze what went wrong before continuing [1].

Advantages of A/B testing are that it produces actionable results, reports actual user behavior, and accurately evaluates impact of specific changes. Disadvantages include a lack of insight into why a user made a certain decision, optimization for one small detail at a time, and the need for a large user base to detect an effect.

Eye tracking is a technology that records the precise point at which a user's gaze is fixated on a screen. This method has more recently gained acceptance among the HCI community due to its ability to measure and predict user actions [13]. The technique involves measuring either where the eye is focused or the motion of the eye as a user views an interface [39]. This technique introduces a whole new layer of understanding user perception since it gives objective insight into exactly what users see and don't see on a screen. Eye tracking devices can measure where the user is looking, how long they are looking there for, and their changes in focus from one item to another.

Advantages of eye tracking include indication of exactly where users are looking, insight into user actions of fixation, insight into pathway of movement within a page, insight into the distinction between gaze patterns of various user groups. Disadvantages are that there is no guarantee that a user saw something conscientiously since they may have spaced out, no guarantee that a user did not see something since peripheral vision is not captured, the necessity for many users, problems testing with certain pupils, process can be costly, and can seem invasive to users [13].

3.3 Measuring Usability

In usability testing, success measures are collected quantitatively and qualitatively.

Performance measures are a user's actions that we can measure quantitatively. They include counts of actions, measures of time, and behaviors we can typically see. Most performance measures are counted through careful observation and do not rely on the observer's judgment. For instances when there is only one way to accomplish a task, the user's choice can be reduced to right or wrong. There are a few performance measures in which the observer needs to decipher a user behavior and decide the right call. Some examples of this would be the number of signs of frustration, confusion, or satisfaction expressed by the user. Performance measures are typically collected through timing or data-logging software. Logging programs can help keep track of things like time and whether or not a user takes a particular action. The act of logging data can either be automated by a *logging program* or be logged by an observer called the *data recorder*.

Subjective measures are a user's opinions, perceptions, and judgments of the product. They can be either quantitative or qualitative. Quantitative subjective measures include ratings and preferences–for example, giving users a 5-point Likert scale to rate how easy the product is to use helps quantify a subjective opinion that we can then compare among users. Qualitative subjective measures come from verbal reasoning and spontaneous comments from the user [14]. Subjective measures can be extracted from questionnaires and asking the user to think out loud.

For the purpose of my research, I will be using both performance measures and subjective measures. The variables measured, including whether they are performance or subjective measures, are summarized in Table 1, while the rationale for all variables is developed in Chapter 5.

3.4 Negative and Positive Behavior Metrics

The goals of usability testing are to find the problems and weaknesses in the product so that we can improve upon the product before it goes out to users. The initial data collected about a product is focused on counting indicators that your goals are not being met. Thus, it is typical to collect data on time, errors, and frustrations in the initial iterations of the product [14]. Low rates would indicate that users are working well with the product.

In addition to tracking unwanted behavior, positive findings can also be an indication that the product's goal is being met. Positive performance measures can be useful in determining a product's usability and future direction. For example, expressions of satisfaction, especially common themes among users, can help to create a more informed product vision. Another way to measure positive performance is to measure the use of a certain features within your product. For example, we may want to measure the number of times users go to the help desk of your product. The goal for the metric may be that users go to the help desk but then promptly return to the main usage once they have figured out how to solve their problem. In this case, clicking the help desk once would be a positive indicator of positive performance, and revisiting to the help desk frequently may be be a positive indicator of negative performance [14].

4 Prototyping

4.1 Iteration 0.5

To get an initial idea of what the Stories medium would look like and feel like, I used Snap Publisher and created a rapid prototype to test out the concept. I created a tap-by-tap talk based off the paper [50] that inspired this research, utilizing the headings of the paper as a guideline for the headings on each screen. Through the creation process I was able to immerse myself in what the process may be like for the crowd volunteers. I then tested the output with three users for first impressions. I noted some takeaways in (a) the process of making the talk myself and (b) testing the flow with a few users.

Figure 9: Initial rapid prototype of a sample slide created using Snap Publisher

4.1.1 Pilot User Feedback

On the producer end, the biggest challenge of making the talk was the reduced screen real estate: with the smaller screen size of this format, there will be a new challenge of enforcing a character limit. The research paper content will need to be distilled down to a minimum and highlight only the necessary research aspects.

On the consumption end, all three users tapped through the talks without instruction and gave positive feedback about the animated elements. Two users stated they would have liked to see some visual media: the initial prototype did not include any visually rich content aside from the background. The Stories medium by nature is visually rich and narrative, and it would be important to incorporate forms of multi-media from the research papers whenever possible in order to keep users engaged.

4.2 Iteration 1.0

Using AMP Stories and AMP HTML, I built and tested two ways of telling the "research story," and had users tap through and react to both styles in randomized order.

4.2.1 Lessons Learned

Based on my feedback from version 0.5, I made some general changes to the style and content of version 1.0.

1. **More visual, less verbal:** Added figures and graphs from the research paper to explain the research content more visually.

2. **Relevant backgrounds:** Background images are selected to complement the content of the slide and aid in consumption. As it's the first snapshot the reader sees even before the text appears in some cases, it helps them contextualize the content they are reading.

3. **Next steps slide:** Last slide added to redirect readers to learn more about the research topic and how they can get involved in the Stanford Scholar initiative.

The changes above applied to both the academic and journalistic formats. Otherwise, the experimental style of the journalistic format was meant to test differences in reading experience for the user.

4.2.2 Style I: Academic

My first approach as shown in Figure 10 was more traditional: it was the most loyal to the breakdown of the Stanford Scholar videos and computer science research paper structure, with the current state of the problem, research question, main contributions, findings, real world apps, etc.

4.2.3 Style II: Journalistic

After studying the design of the stories on Snapchat's Discover tab, and watching Two Minute Papers, I wanted to test out an experimental format that was more journalistic. This format was inspired by the narrative nature of AMP Stories, and was broken up more like a news story, as shown in Figure 11.

4.3 Challenges

There were three consistent challenges that users ran into when interacting with both formats of version 1.0.

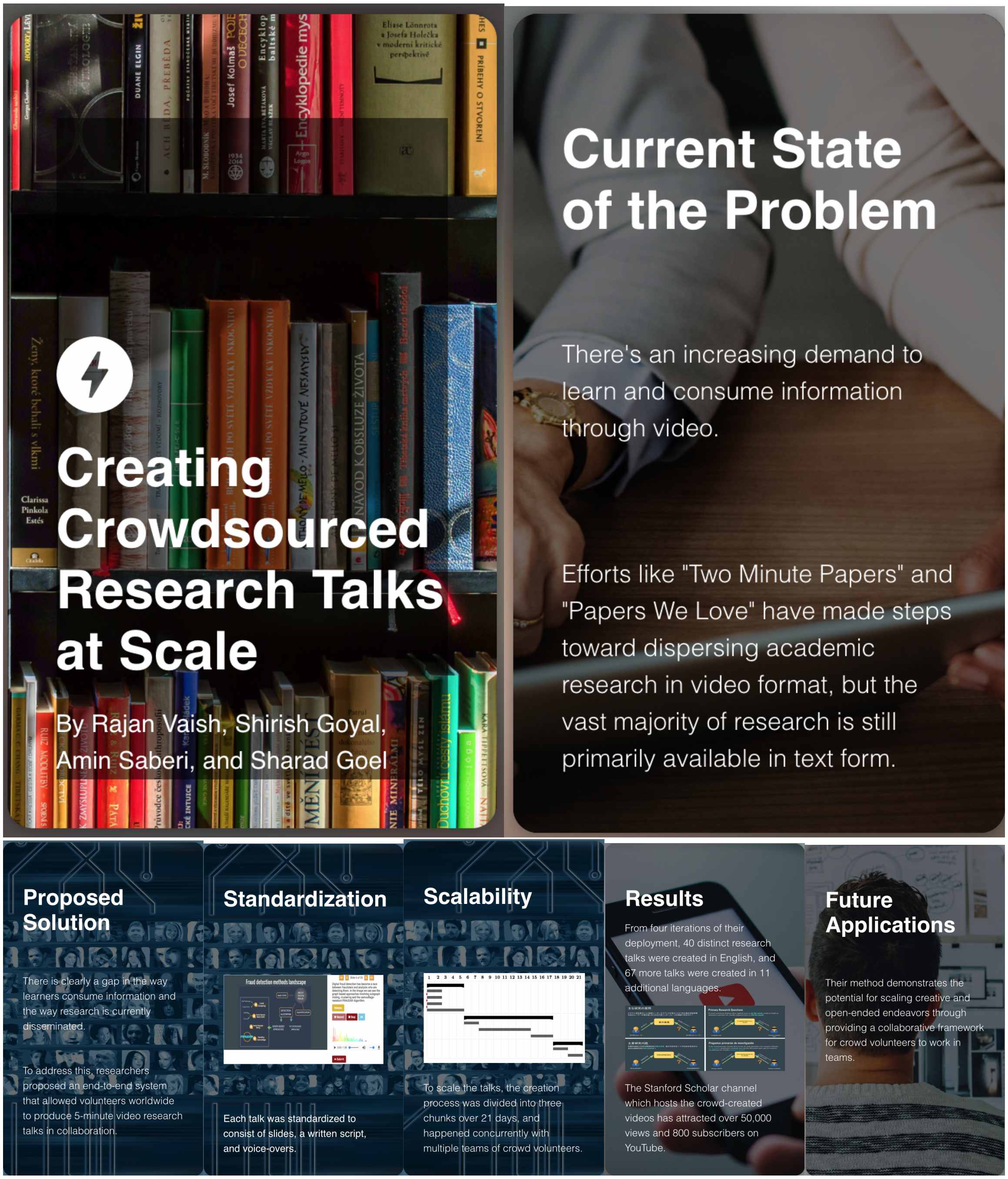


Figure 10: Flow of academic format.

1. **Video interaction:** No users clicked on the video presented to them first. One user said they were focused on the text while another was expecting autoplay: "Should I click on this? I was expecting it to just play since videos do that in the Discover tab."

2. **Applied vocabulary:** A large majority of readers mispronounced the word

disseminated, one pointing out that they understood the message but were not sure if this would be approachable vocabulary for everyone outside a technical field.

3. **Faces in the background:** 15% of users pointed out that the faces in the background distracted from the text. Based on eye-tracking data [24], users tend to be drawn to faces when asked to look at web pages. This would be important to indicate in the best practices for choices of background.

4.4 Iteration 2.0

Finally, I used Heroku and AMP Stories to build a high-fidelity prototype to demonstrate the full experience on mobile devices and test for interaction.

4.4.1 Lessons Learned

Based on my feedback from version 1.0, I made the following changes to the overall concept of version 2.0.

1. **Animation:** After receiving positive reactions about text and media animation and movement, incorporate some form of dynamic movement in every slide using AMP HTML animations.
2. **Minimalism + Focus:** Based on user feedback in version 1.0, avoid mixing a lot of media with text. The media should be briefly explained but should remain as the focus of the slide.
3. **Autoplay:** Set embedded videos to autoplay to fit the fast pace of the Stories medium and expectations.
4. **Journalistic Leaning:** Users indicated that both formats in Iteration 1.0 were readable and understandable. The academic format was clear in what it was trying to convey while the narrative format was more of a pleasant experience.

4.5 Generalizing to a Template

After three iterations of exploring how best to convey the research talks in the Stories format, I created a generalized template that would serve as a guide in building the Stories research talks, as outlined in Appendix C.

5 Experiment

5.1 Research Questions

Since there are two sides to the video-making process–the creation experience and the consumption experience–there are a few ways to measure success of the two mediums.

One approach would be to look at the crowd volunteer experience of making the talks through a comparative study of the *process* of making the videos versus stories talks from the creator perspective. Another way is to focus on the end learner experience of consuming the research through video vs. through stories. For the purpose of this research, I will constrain the problem to the consumption side since I am looking to answer a more objective question related to how individuals learn information through these distinctive mediums without introducing the convolution of collaboration which has already been shown to enhance engagement and critical thinking [7, 21].

To extend the literature on the Snapchat and Instagram Stories format in the context of education and to explore new methods to consume educational content more generally, I conducted a user study with 22 participants. My study explored three key research questions comparing the Stories versus Video medium on the *consumer* end of the experience:

1. *Educational value.* How well did participants learn the key contributions of the paper through Stories versus through Video?

2. *Satisfaction.* Does the user prefer to experience the research talks through a particular medium?

3. *Efficiency.* Does a certain medium allow the user to finish the task quicker?

Since the speed of consumption is now in control of the user, I measured the time it takes for the user to experience the talk through a story versus a video. Watching videos is a passive experience [30]. Tapping through a collection of pages may be a more active way to consume content which would contribute to greater takeaways in learning.

5.2 Overview

To understand if the Stories medium is better for academic research consumption than Video, I designed a comparative study to (1) evaluate the educational value of each medium, (2) measure the efficiency of each format in consuming the academic research content, and (3) test the user's satisfaction and preference for a particular format based on their expressions of satisfaction and post-task preference indication.

My definition of educational value is further broken down into self-reported measures of (1) attention, (2) post-task understanding of key contributions of the paper, (3) post-task improvement of paper's subject matter. I take into account the potential confounding effect—as well as others, described in the Study Design—of a participant's previous knowledge of the paper topic by asking them to rate their own expertise on the paper topic on a scale before going into the study.

My measure of efficiency is further broken down into (1) the *time* it takes for the user to experience the research paper in each medium, (2) the *number of errors* the user makes in each medium that may slow down their rate of consumption, and (3) a post-task self-reported measure of how *quickly* the user felt they were able to grasp the paper's key concepts in each medium.

Lastly, I break down a user's preferences into (1) their *expressions of satisfaction* throughout the study, (2) a self-reported measure of how *boring* they found each medium, (3) a self-reported measure of how *compelling* they found each medium, and (4) a self-reported indication of which format was more *pleasant to learn from.*

5.3 Study Design

Structure. The study was divided into a pre-task survey, four tasks on mobile, and a post-task survey as outlined in Figure 12, and was designed to take 20-30 minutes.

Both performance measures and subjective measures were used to examine the research questions. Screen recording available through iOS on iPhone was used to examine the user's interaction with the interface during the four tasks on a mobile device. The screen recording collected information about the performance measures such as time to complete each task, and number of user errors for each task. The performance measure of expressions of satisfaction was judged subjectively throughout the study based on the number of times they laughed or made a comment indicating satisfaction with the interface.

A pre-task questionnaire was used to better understand the user's background and familiarity with the technology, medium, and academic research being presented. This information was to meant to help understand how expertise in a topic affects how well a user understood the research content.

The four tasks on mobile included two research papers presented in both Stories and video mediums on a mobile device. The motivation behind showing the user two different papers in reversed format was so that the user's preferences were not biased by the improvement of their understanding of the content just because they had seen it a second time. Since the user was not being tested on their understanding of the content, and rather exploring their preference for a particular format, the content of one research paper experience within the two mediums was designed to be exactly the same, in order to ensure that the user could naturally develop their preference for a certain format. Although the order in which the academic papers was shown was irrelevant, the order of the mediums was kept inverted for each paper. In other words, it did not matter if academic paper #2 was shown first, so long as it was followed by academic paper #2 in the other medium later in the study, and then in reversed order for academic paper #1.

The post-task questionnaire was used to further examine the research questions through subjective measures of educational value, preference, and satisfaction with each interface. This information was used to explain the user's preferences and how the user understood the research through the distinct interfaces.

Considerations and confounding effects. There were a few careful considerations in the study design aimed to reduce anticipated confounding effects. The first was the choice to use a within-subjects experiment over a between-subjects experiment for this work. One of the advantages of using a within-subjects experiment—sometimes referred to as repeated measures design—is that it takes fewer subjects to get the same amount of data. Another advantage of repeated measures design is that it is better at detecting differences. When trying to detect differences, we are doing so “against a backdrop of natural variation, measurement error, and general noise” [51]. Therefore, any way to reduce variance within measurements will provide more accuracy regarding the factor being measured. Due to the absence of individual differences, there is generally less variance in within-subjects studies. The disadvantage of within-subjects design is that it may introduce carryover effects. Since each user performs the task in both mediums, the order in which they perform the tasks may cause the user to perform differently [28]. In the context of my study, the user would experience the same academic research content back to back with the only change being the change in medium. The confounding effect here would be the user thinking they understood the material better through the second medium they saw the content with, not because of the medium, but because it was their second pass through the content. Given the deep ties the research questions have with educational value, it was crucial to mitigate the potential of this confound. For this reason, there was a second piece of content added to the study which would be showed in opposite order of format, as detailed in 2a–d in Figure 12. The addition of a second research paper will allow participants to have clearer judgment on their preferences and abilities to learn from a particular medium since they will have experienced the content from both mediums in opposite orders.

Another confounding variable considered was prior exposure to the research content. Participants who were more familiar with the topics of the research papers such as AI or computer vision may have been more comfortable with clicking through the Stories format more quickly, whereas the video consumption is more or less fixed. To control for this, I (1) tried to recruit computer science majors and students interested in the paper topics presented and (2) added a question in the pre-task survey that asked participants to rate their familiarity with the paper topics before performing the four tasks.

The question design within the survey was another important consideration in the study. In building surveys, it is important to avoid leading questions that encourage certain responses [47, 40]. Before administering my survey, I had two peers review it to be more sure that participants would be able to freely agree and disagree with my proposed questions. I also had reviewers give me feedback about whether or not the survey questions encouraged certain responses for certain mediums. I added three free-form responses so that participants would be able to justify their preferences with less restriction.

5.4 Participants

I recruited a total of 22 participants through announcements after computer science classes and public meetings. The announcements sought to motivate potential participants with the opportunity to learn more about cutting-edge research in computer science. Participants who expressed interest in participating in the study were contacted through email to set up a time to do the study. Only adults over the age of 18 were recruited for the study. Participants were not paid or otherwise compensated, and their participation was completely voluntary.

The majority of participants were students from the five Claremont Colleges who studied computer science. The next most common fields of study included economics, PPE (Philosophy, Politics, & Economics), and Mathematics, as shown in Figure 14. Most participants had been exposed to the topics of the research papers beforehand, and were asked to rate their familiarity with the topic in a pre-task survey as this was an anticipated confound effect. The percentage of women participants was 36%. Participants typically had moderate technical expertise in the paper topics they were exposed to in the study, as illustrated in Figure 15. Based on self-reports, the mean expertise was 5.4 and the median expertise was 6 on a scale from 1 to 10, with 10 indicating the highest expertise. Most participants indicated that they had high familiarity with the tappable Stories medium on mobile, as shown in Figure 16. Based on a self-reports, 86% of participants indicated that they were either very familiar or extremely familiar with the Stories medium. This supports prior literature findings regarding the reach of platforms with Stories content for the teenage and young adult demographic, as shown in Figure 13.

5.5 Privacy

This work was submitted to the Institutional Review Board at Claremont McKenna College and received approval on April 2, 2019.

Identifying information. The extent of the identifying information participants were asked to indicate was whether they were a student or faculty member, their area of expertise or study, and an initial self-reported rating expertise of the paper topics in the study. This information was abstracted in the results to ensure that no participant could be reverse-identified. The information was not included in the final analysis. Additionally, the screen recording of user behavior during the study was kept private and was accessed only by the researcher.

Consent. All participants were asked to provide written consent prior to beginning the study. Participants acknowledged their desire to participate in the study by signing a copy of the informed consent form and confirming their understanding that they may withdraw their participation at any time, for any reason, without penalty. Participants also gained a clear understanding of the purpose of the research as described through the consent form. Additional verbal confirmation of consent prior to the interview gave all participants the opportunity to address questions or concerns. Comments made during the study were also transcribed, and participants were asked to consent to their statements being included in the final publication if applicable.

5.6 Procedure

Once the participants signed a consent form explaining the screen recording, what the mobile experiences will entail, and what type of data will be collected, they went through and filled out a pre-task survey. The information collected in the pre-task is for the purpose of gaining a better understanding of participants' backgrounds and familiarity with the Stories format and the paper topics. After completion of the pre-task survey, I explained the order of the tasks to them and what they would be watching. For the Stories experiences, participants would tap through the experience at their own pace, and the recording of the total time would be logged from the screen recording. It was suggested that the participants go through the video experiences as they would if they were learning on their own, and to rewind and pause as needed. Because of this, the video lengths were not fixed to their set play length. The total time taken to watch the video was logged through a screen recording. After the four tasks, the participants moved on to complete a post-task survey about their experiences.

6 Results

The 22 participants provided 22 observations of 14 different variables. The variables were divided into two data sets: the observed performance measures from the four tasks and the subjective measures from the survey results. Collectively, each observation was a 14-tuple.

Variable	Type	Performance or Subjective	Target Research Question	Location in study	Self-report
Time to consume content	numerical	P	Efficiency	During task	No
Attention	categorical	S	Educational value	Post-task	Yes
Expressions of satisfaction	numerical	P	Satisfaction	During task	No
Understanding of key contributions	ordinal	S	Educational value	Post-task	Yes
Improvement in knowledge of paper's subject matter	ordinal	S	Educational value	Post-task	Yes
Pace of comprehension	ordinal	S	Efficiency	Post-task	Yes
Compelling rating	ordinal	S	Satisfaction	Post-task	Yes
Boring rating	ordinal	S	Satisfaction	Post-task	Yes
Pleasant to learn from	categorical	S	Satisfaction	Post-task	Yes
Errors made	numerical	P	Satisfaction	During task	No
Familiarity with the "Stories" medium	ordinal	S	Educational value	Pre-task	Yes
Familiarity with the paper topics	ordinal	S	Educational value	Pre-task	Yes
Student or faculty member	categorical	S	Educational value	Pre-task	Yes
Area of study	categorical	S	Educational value	Pre-task	Yes

Table 1: Summary of variables and how they were measured in the study.

Factors. An important factor in the study is the medium in which the research content was presented. The levels of the factor are the medium type, i.e., Stories and Video. So this is a categorical variable. Since both levels of the factor were tested with each participant, the study was a within-subjects experiment [51].

Response. The response variables of the experiment included a mix of observed performance measures and some subjective measures. The variables along with their types and descriptions are presented in Table 1.

6.1 Overview

In the study, 22 users interacted with the two different research papers through both the Stories and Video mediums to make for two-factor within-subjects experiment. All statistical tests for analysis were performed in R.

The statistical tests available are presented in Table 2. These include the paired-samples t-test and the Wilcoxon signed-rank test. The Wilcoxon signed-rank test is the nonparametric equivalent of the parametric paired-samples t-test. Many parametric tests require that the three ANOVA assumptions of normality, independence, and homoscedasticity are met [51]. Thus, the type of analysis used will depend on the variable and its adherence to the ANOVA assumptions.

Factors	Levels	Between or Within	Parametric Tests	Nonparametric Tests
1	2	B	Independent-samples *t*-test	Mann-Whitney *U* test
1	>2	B	One-way ANOVA	Kruskal-Wallis test
1	2	W	Paired-samples *t*-test	Wilcoxon signed-rank test
1	>2	W	One-way repeated measures ANOVA	Friedman test
>1	≥2	B	Factorial ANOVA Linear Models (LM)	Aligned Rank Transform (ART) Generalized Linear Models (GLM)
>1	≥2	W	Factorial repeated measures ANOVA Linear Mixed Models (LMM)	Aligned Rank Transform (ART) Generalized Linear Mixed Models (GLMM)

Table 2: Summary of analyses of variance by experiment type.

6.1.1 Testing for ANOVA Assumptions

We must test our numerical variables for the three ANOVA assumptions of independence, normality, and homoscedasticity so that we can begin to decipher if we can use a parametric test. The assumption of independence is met through sound experiment design, and more specifically if (a) each user is sampled independently of every other user, and (b) measures on a user are independent of measures on every other user [51]. Independence applies to all variables measured in this study, as verified through the experiment design discussed in 5.3. To test for normality, we look at the data distributions and the Shapiro-Wilk test for normality on the responses and also on the residuals, which are the difference between our observed measures and the predictions of the statistical model we use [51]. Finally, the homoscedasticity assumption can be tested with Levene's test and the Brown-Forsythe test [51]. None of the variables in this study passed all three ANOVA assumptions, and thus parametric testing was not used.

6.2 Numerical Data

Time to consume content. Since time is a continuous variable, a kernel density plot was used to analyze the distribution in order to better understand behavior. The kernel density indicates that the data for the time variable for Stories and Video is unimodal with some outliers. As shown in Figure 17, time for Stories in both papers tended to have a large global maximum around 160 seconds. Time for Video tended to have more distinct global maximums, with a maximum around 211 seconds for Paper #1 and 215 seconds for Paper #2. Unsurprisingly, these peaks for Video were around the times of the video lengths (212 seconds for Paper #1 and 213 seconds for Paper #2).

To investigate the distribution further and begin to look at the differences in time for both mediums, we can look at the boxplots in Figure 18. The distribution of Stories was greater than Video in both papers, which makes sense because Stories are self-paced and Video is more or less a set pace for the user. The boxplots help visually suggest that there is quite a significant difference between the time it takes to consume the content through each medium. Faster readers are likely the ones to make up the lower portion of the distribution of Stories, but even in the Stories' worst case for paper #2, it takes the slowest reader at least the same amount of time to consume the academic content on average. There were a few outliers in both mediums who took longer to go through the experiences, but overall, Stories seemed to be a faster way to experience the exact same content.

To test for normality more confidently, we can perform the Shapiro-Wilk test. For this test, the low p-value results, as outlined in Table 3, suggest that the distributions for time in both mediums stray too far from a normal distribution. The p-value for Stories in paper #1 is just above 0.05. The p-values for Stories and Video for the rest of the papers are far below 0.05. This is also the case with the residuals as shown in Table X. Since we cannot say that time has a normal distribution, we move on to nonparametric tests.

The results of the Wilcoxon signed-rank test for time are summarized in Table 5. Both p-values were significant, which suggests that it takes less time to consume the same information in Stories compared to Video.

For 94% of users, Stories was a faster way to consume the same research content. The relationship between time for Stories versus time for Video can be seen in the scatterplots in Figure 20. Each point in the scatterplot represents the time that the same user took to consume content in each medium. The line y=x represents a theoretical scenario where Stories and Video take the same amount of time. The 83 points underneath the line in Figure 19 and Figure 20 represent all the users who consumed the same content faster through the Stories medium. The 5 points above the line represent the users who consumed the content faster in the Video medium.

Expressions of satisfaction. The distribution for expressions of satisfaction is shown in the histogram in Figure 21. Expressions of satisfaction and Errors made both seemed to resemble Poisson distributions as suggested by the results of the goodness of fit statistic in R in Table 7. However, these tests should be taken with a grain of

	Stories #1	**Video #1**	**Stories #2**	**Video #2**
p-value	0.070	9.9e-08	0.0050	2.8e-08
W	0.92	0.49	0.86	0.43

Table 3: Shapiro-Wilk normality test results for time separated by medium and paper.

	Paper #1	**Paper #2**
p-value	0.033	0.0046
W	0.90	0.85

Table 4: Shapiro-Wilk normality test results for residuals of time.

Paper	**p-value**
1	9.5e-07
2	0.0011

Table 5: Results of Wilcoxon signed-rank test for time.

Variable	**p-value**
Expressions of satisfaction	0.35
Errors made	0.078

Table 6: Results of Wilcoxon signed-rank test for count data.

salt since the data set is relatively small.

Medium	**Chi-square p-value**
Stories	0.61
Video	0.40

Table 7: Poisson goodness-of-fit test results for expressions of satisfaction.

The results of the Wilcoxon signed-rank test for expressions of satisfaction were not significant. This tells us that we don't have strong statistical evidence that there was a difference between between satisfaction expressed for the Stories medium over the Video medium.

Errors made. There were a few variables that did not satisfy the assumptions for normality. Errors and count data generally are a very common example of why nonparametric tests are necessary in analyses of variance since they stray from Gaussian distributions and are often Poisson distributed. In the case of *expressions of satisfaction* (Figure 21) and *errors made* (Figure 22), both distributions resembled a zero-inflated Poisson distribution [32]), as is typical of count data.

The results of the Wilcoxon signed-rank test for errors made were also not significant, and so there was weak statistical evidence between the distinction of amount of errors users made on either medium. Users made significantly more errors when using the Stories medium. The withstanding errors made in Stories may be associated with the learning curve of adapting to a new medium.

The results of the Wilcoxon signed-rank test for all the count data are summarized in Table 6.

6.3 Ordinal Data

Although using parametric tests is valid in some situations for Likert scale items [19, 33], these types of responses typically do not satisfy the conditions for ANOVA [51, 26]. In the case of my Likert scale variables, although the data is ordinal and there is an equivalent *numerical* difference between 4 to 5 and 3 to 4, the *conceptual* difference between "somewhat familiar" and "not familiar at all" does not necessarily scale across all responses. The same issue applies to all ordinal variables in the survey. For this reason, the 7 ordinal variables from this study will be analyzed using a nonparametric test.

Understanding of key contributions. The responses for contribution understanding were set on a scale of 1-5 of how well the user thought they learned the key contributions of the paper after experiencing the content in each medium, with 1 representing "Not well at all" and 5 representing "Extremely well." Stories had a mean of 4.1 (median = 4, SD = 0.64), suggesting that users generally learned the key contributions very well. Video had a slightly lower mean of 3.8 (median = 4, SD = 0.61). The distribution of understanding grouped by medium is shown in Figure 23. The means are similar for both mediums, but the interquartile range (IQR) for Stories is a bit higher, ranging from 4-5, while the IQR for Video ranges from 3-4. This suggests that there is quite a significant difference between the time it takes to consume the content through each medium.

The results of the Wilcoxon signed-rank test for this variable are statistically significant, as summarized in Table 8, indicating that there were detectable differences between mediums in regards to understanding key contributions in Stories over Video.

Improvement in knowledge of paper's subject matter. The responses for knowledge improvement were set on a scale that measured how much the user felt their knowledge of the paper topic had improved from experiencing the research talk through a particular medium, with 1 representing strong disagreement and 5 representing strong agreement. Stories and Video both showed strong agreement with a mean of 4.5 (median = 5). The standard deviation was 0.67 for Stories and 0.60 for Video. The histograms (Figure 24) and the stark p-value of 1 resulting from the Wilcoxon signed-rank test (Table 8) emphasized that there was no distinguishable difference between mediums in relation to improvement in the user's knowledge about the subject matter.

Pace of comprehension. Users expressed how quickly they were able to grasp the key concepts of the paper on a Likert scale of 1 to 5. The scale was based on agreement regarding their ability to grasp the concepts quickly through a certain medium, with 1 being "Strongly disagree" to 5 being "Strongly agree." The mean for Stories was high (4.7, median = 5, SD = 0.57), and was lower for Video (3.5, median = 4, SD = 1.1), which suggests that users are able to comprehend the key concepts of a paper quicker through the Stories medium.

The histogram in Figure 25 further illustrates that Stories might be favorable to Video in terms of a user's pace of comprehension, with 73% participants indicating "Strongly Agree" to Stories for their response to the question "I was able to grasp the key concepts of the paper quickly."

The Wilcoxon signed-rank test helps to further examine the data: the resulting p-value is statistically significant, suggesting that there is a detectable difference between Stories and Video in terms of the pace of comprehension. The results of this subjective measure of efficiency complement the earlier results of Time, a performance measure of efficiency, and together suggest that Stories are a quicker way to consume the same content.

Compelling rating. Users expressed if they found a particular medium compelling on a Likert scale of 1-5, with 1 representing strong disagreement with being compelling and 5 representing strong agreement with the statement. Users seemed to find Stories slightly more compelling, with an average rating of 4.5 (median = 5, SD = 0.67). The mean rating for Video was 4.0 (median = 4, SD = 1). These results are illustrated visually in the histogram in Figure 26, suggesting a slight preference to Stories in terms of how compelling they were.

However, according to the insignificant p-value that resulted from the Wilcoxon signed-rank test for this variable, as shown in Table 8, there was no strong statistical evidence that would suggest a difference between how compelling users found Stories compared to Video.

Boring rating. Like the compelling rating, users expressed if they found a particular medium boring on a Likert scale of 1-5, where 1 indicated strong disagreement with the medium being boring and 5 indicated strong agreement with the statement. There was slightly more disagreement with the statement that Stories were boring (mean = 2.2, median = 2, SD = 0.85) as compared to Video (mean = 2.9, median = 3, SD = 1.2), and this could be further compared visually by the histograms in Figure 27. There was also a larger spread of opinions regarding how boring Video was, as illustrated in the histogram in Figure 27.

The results of the Wilcoxon signed-rank test (Table 8) were significant, which provided stronger evidence that Video was perceived as more boring in comparison to Stories.

Variable	**p-value**
Understanding of key contributions	0.038
Improvement in knowledge of paper's subject matter	1
Pace of comprehension	0.00032
Compelling rating	0.30
Boring rating	0.037

Table 8: Results of Wilcoxon signed-rank test for all ordinal variables.

6.4 Categorical Data

Categorical variables from the study–including attention, pleasant to learn from, area of study, and student or faculty indication–did not require analysis further than a look at their distributions and relationships with other variables, and were excluded in the testing for ANOVA assumptions. Findings from the categorical variables were complemented by the qualitative feedback gathered from the free-form responses in the survey, in which participants were asked to justify their responses.

Attention. When asked which medium kept their attention better, 72% of users indicated Stories, 23% indicated Video, and 5% had no preference, as shown in Figure 28. When asked to justify their responses, there were a few common themes that stood out for each medium. All 16 user responses from those who felt that Stories kept their attention better is referenced in Table 9. From examining the Stories feedback more closely, there were a few key themes.

- **Control of pace**: Among the array of feedback for Stories, one of the themes that came up repeatedly was how Stories allowed users to go at their own pace. 11 out of 16 users directly referenced pace, through phrases like "it was nice to move on to the next section with a tap at my own pace," "Easier to go at my own pace," and "I felt more in control of the pace" suggested that users appreciated being able to control the pace of the content through the Stories which the video medium didn't allow them to do. Stories allowed users to process the information at their own pace, which was often quicker than the time they consumed the content in the video, as suggested by the results for *Time to consume content*.

- **Control of content**: Aside from being in control of the pace of the experience, participants also stated that they felt as if they had more choice in the content they wanted to focus on, and were able to "skip over content [they] deemed unnecessary." One user talked about how the ability to control the pace in turn allowed them to control the content they wanted to view: "Because I had some ability to control the pace of the information, I could skip over parts I was done learning about to focus only on the parts that interested me the most." Another user suggested that Stories gave them more precise control of content: "When learning through video, if I miss a concept (by a lack of congruency in audio and video for example) then I have to rewind without precision leading me to tediously relearn or spend needless time." Although the concept of "bite-sized content" has been explored in the space of journalism, the primary industry using Stories, its potential in education fits into the paradigm of microtasking and microcontent [48, 9] since the Stories medium provides the user with actionable steps to consume the research paper.

- **Ease of navigation**: A few users made statements referencing ease of navigation, saying that the Stories were "easy to navigate and condensed the information step by step," and that they "liked the interactive component" whereas navigation in Video involved having to "rewind without precision." Even a user who chose Video stated in their response that they found it "easier using the snap version to return to where [they were] and pick back up where as with the video [they were] less likely to do that." Ease of navigation supports the *learnability* of the interface, one of the components of usability as defined by Nielsen [37].

- **Active participation**: Another theme that arose in the Stories feedback was active absorption of information that resulted from the tap-through feature. Users said that the medium allowed them to "actively read and advance the information," "actively participate" by tapping, and ensure that they were "actively involved in the experience." Although the content shown was identical, the type of learning that was happening may have been different in the Stories medium. There was a clear distinction in feedback from users who preferred Stories since they were an "active" experience compared to the feedback about liking the videos because of the passivity of the experience. This distinction suggests that their preference may be attributed to learning style. Active learning places more responsibility on the learner by involving them in an activity, while passive learning involves receiving information and internalizing it [30]. Users who preferred "taking [their] time when reading through information" seemed to prefer the Stories medium since they had to actively advance the information themselves.

- **Reduced distraction**: A few users talked about how it was "harder to zone out" in the Stories medium, and that there was "reduced distraction." This theme of reduced distraction in Stories went along with another user's critique of the Video medium being overwhelming: "In the videos I was often overwhelmed by needing to both read the text shown in the videos while also listening to the speaker, and so I was notable to grasp as much information or at least was not able to grasp is as deeply as I was able to in the stories." The reduced distraction of the Stories medium may allow users to learn content in less time, while the microcontent format of Stories allows users to learn more content, together leading to more efficient learning.

The feedback from the 23% of users who indicated that they thought video kept their attention better is listed in Table 10. The common pattern from the feedback for video was an appreciation for voice and sound that was part of the video experience. Some users particularly enjoyed the voice of the speaker, and how it directed their attention to what was happening stating that they "found his voice entertaining because it wasn't monotone," and that since the "voice was engaging, [their] attention was directed more automatically to where it should be." Users who preferred video in terms of attention also expressed a preference for a passive learning style with phrases like "it is easier to watch a video and listen to a voice than read," and that it is "easier to pay attention to the video and listen to the commentary instead of the stories where [they] had to direct attention to words and then to the video in the background." Enjoying the speaker's voice may have confounded some user's decisions if they paid attention better because they found the speaker's voice charming rather than paying attention to any voice. A future consideration may involve incorporating an audio component to the Stories. It is currently possible to incorporate sound to Stories, but it will also be important to consider that the rate at which the user will be able to tap through the content may then be bound by the length of the audio for each slide.

Although the potential confounding effect of preferring a particular medium over the other because of the order it was viewed in was addressed in the design of the experiment, the one user who had no preference for which medium kept their attention better claimed it was because "it was hard to separate the effects of the medium from the effects of viewing the same information again."

Pleasant to learn from. Users were also asked which medium they found more pleasant to learn from. These responses were not necessarily in line with the user's previous preferences for which format kept their attention better, with 50% of users stating that they found the Video medium more pleasant to learn from, 41% indicating Stories, and 9% indicating no preference. Common patterns that emerged from the analysis of the Video feedback, as referenced in Table 12, included:

- **Narrator as the teacher**: A few users expressed their appreciation for an added human character into the learning experience. They enjoyed being taught by a narrator, saying that "the positive tone in the narrator's voice made it pleasant and exciting," "the script from the narrator helped [them] feel like [they] understood the topic better [since] he gave a little extra context to everything", and that "having a narrator to explain things added more value to the visuals and it added character." Overall, it seemed as if the narrator complemented the content since he added a human and audio component that the Stories lacked.

- **Audio as a complement to visual**: Users particularly expressed that sound was an important aspect of their learning experience, particularly when paired with the visual content. One user explained that she liked the self-paced aspect of Stories, she "liked to have audio and visual components when learning new material." Like the narrator, the component of audio in general seemed to add more value to the content for the learner.

- **Reduced effort**: The users who preferred this medium seemed to prefer listening to reading. They expressed that it was "easier to learn from the video since you do not have to read." One user stated that they "generally prefer listening to reading, unless the content is particularly challenging," while another stated that there was "less effort involved in the overall experience, which led to a pleasant experience." Since the Stories required the user to actively choose when they wanted to move to a new page, this feedback seems to suggest that the pre-set pace of the video experience along with narrative audio made it more pleasant.

Users who felt that Stories were more pleasant to learn from responded with similar reasoning as users who felt that Stories kept their attention better. The same themes of pace ("I'm a slow absorber of information so the stories were better to go at my own pace"), control ("I had more control, feel like I could go about learning at my own pace, my own way"), bite-sized content ("more easily digestible than a single constant stream," "displayed both text and images in a structured way that was pleasant to follow visually") and ease of navigation ("They are easy to navigate and condensed the information step by step") came through in the feedback for why Stories were more pleasant to learn from, as shown in Table 11. One user who preferred Stories reiterated an implicit point made by the preference for video by explicitly stating that they wished Stories included audio. It seemed that users who found Stories more pleasant to learn from thought the format was more digestible and easier to follow.

Some users experienced technical difficulties from connection issues in the Stories format. The background video did not show up for one user in the Stories, who started their feedback with "despite lacking the in-frame experiment footage provided in the video version...." Another user talked about how, "at times the text/description took a while to load which was frustrating because it took longer." Interestingly, both of these users, who were the only users to reference any technical difficulties with the Stories medium, still chose Stories as more pleasant to learn from.

Users who indicated no preference for which medium was more pleasant to learn from (Table 13) did so either because (a) they enjoyed both formats equally pleasant to learn from ("Both the video and story formats both contained the same gifs and images that made the information enjoyable.") or because (b) they found both mediums useful for different scenarios ("If I want to effectively focus then stories are preferential and therefore more pleasant. But if I don't want to physically engage and simply get the gist (for review or relaxed learning), then I prefer video.")

User feedback regarding why Stories kept their attention better
I prefer taking my time when reading through information so it was nice to move on to the next section with a tap at my own pace.
They are easy to navigate and condensed the information step by step.
Easier to go at my own pace
I had more control and was able to absorb the information one fact at a time. When learning through video, if I miss a concept (by a lack of congruency in audio and video for example) then I have to rewind without precision leading me to tediously relearn or spend needless time.
I could go at my own pace
I had to actively participate in the story format by tapping, which allowed me to choose the content that I wanted to focus on and skip over the content that I deemed unnecessary.
I think that the story format kept me much more engaged, for I had to actively read and advance the information at my own pace rather than having this done for me.
In the videos I was often overwhelmed by needing to both read the text shown in the videos while also listening to the speaker, and so I was not able to grasp as much information or at least was not able to grasp is as deeply as I was able to in the stories. I also take a while to process things and I liked the ability to move at my own pace with the stories.
I felt more in control of the pace
I was able to read it at my pace and fully understand what was happening. By having to tap the screen it ensured that I was actively involved in the experience.
I was more engaged because I had to keep clicking through. It was harder to zone out. Also there was less information so easier to absorb.
Because I had some ability to control the pace of the information, I could skip over parts I was done learning about to focus only on the parts that interested me the most.
Stories allow you to read at your own pace whereas videos you have to watch the whole way through.
I liked the interactive component and I could go at my own pace
Because I needed to tap to learn more, in video format it is easier to let your mind wander
The stories allowed me to go at my own pace which reduced distraction.

Table 9: Feedback from users who stated that Stories kept their attention better than Video.

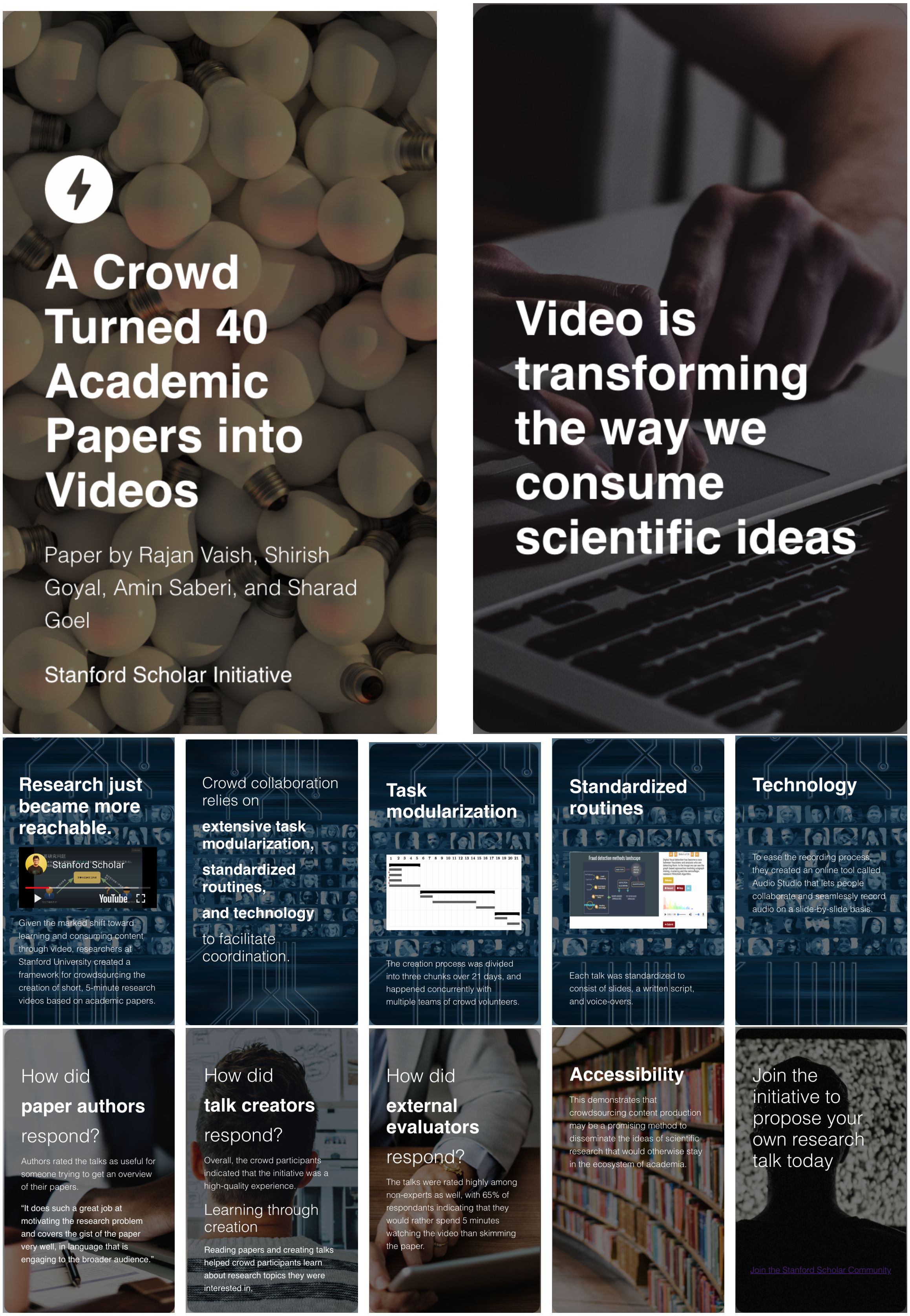


Figure 11: Flow of journalistic format.

1) Pre-task survey [2-5 minutes]
2) Four tasks on mobile
    a) Academic paper #1 presented through Stories tap-through medium [~5 minutes]
    b) Academic paper #1 presented through video medium [~5 minutes]
    c) Academic paper #2 presented through video medium [~5 minutes]
    d) Academic paper #2 presented through Stories tap-through medium [~5 minutes]
3) Post-task survey [5 minutes]

Figure 12: Study structure with estimated times.

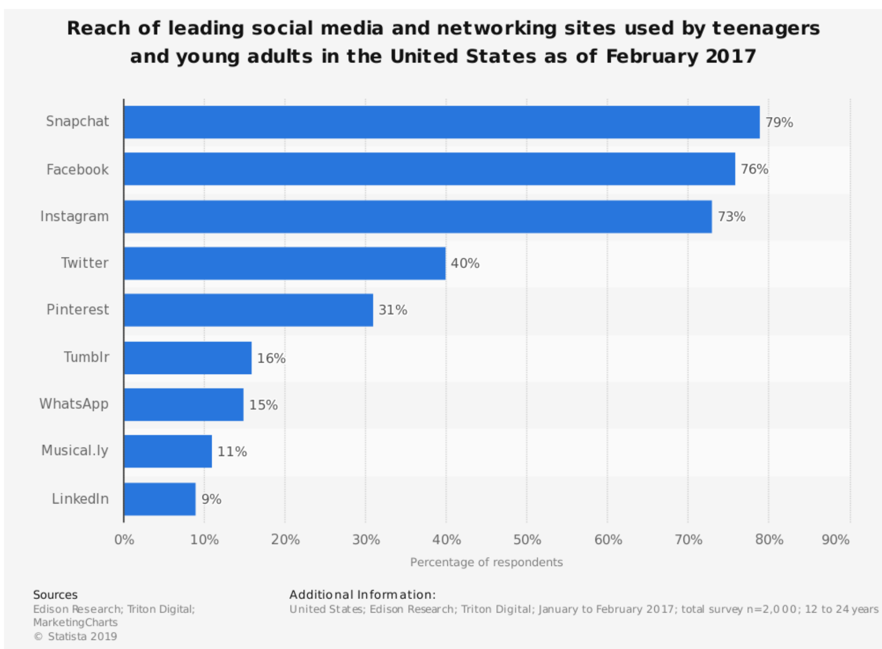


Figure 13: The most popular social media sites used by teenage and young adult internet users in the United States (January to February 2017; total survey n=2000; 12 to 24 years) [45].

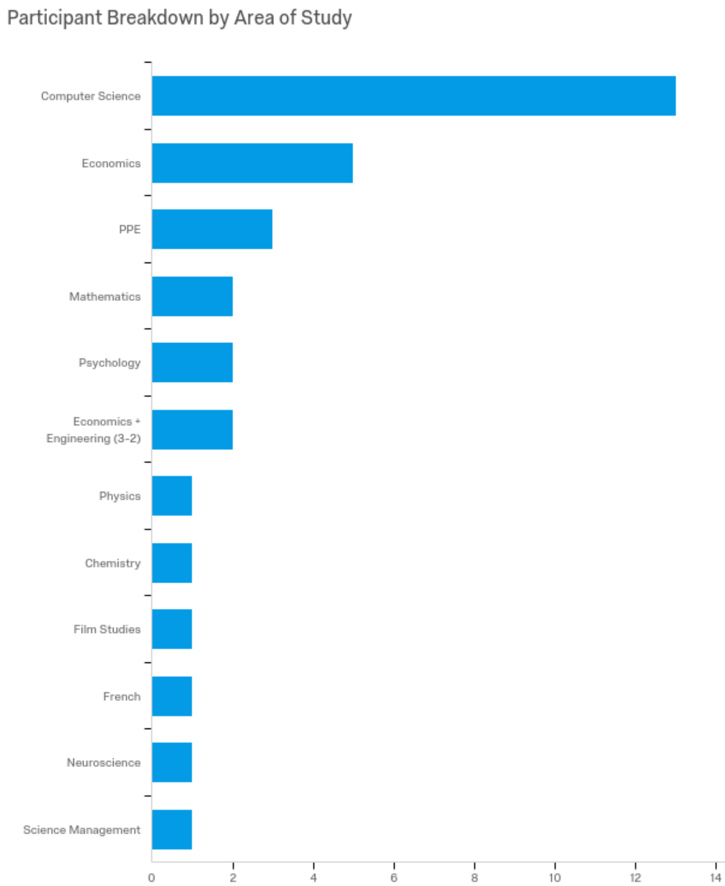


Figure 14: Participant breakdown by area of study.

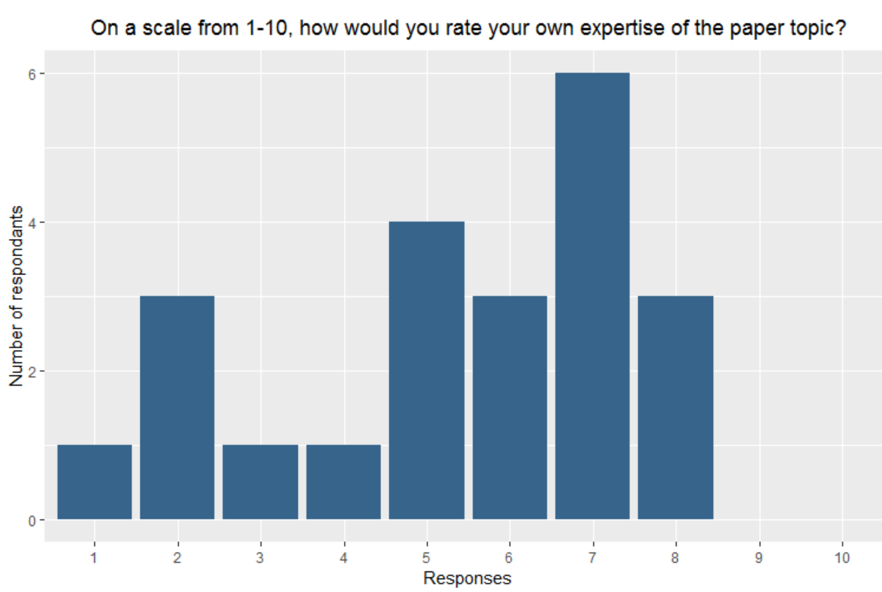


Figure 15: Participant breakdown by familiarity with paper topics presented.

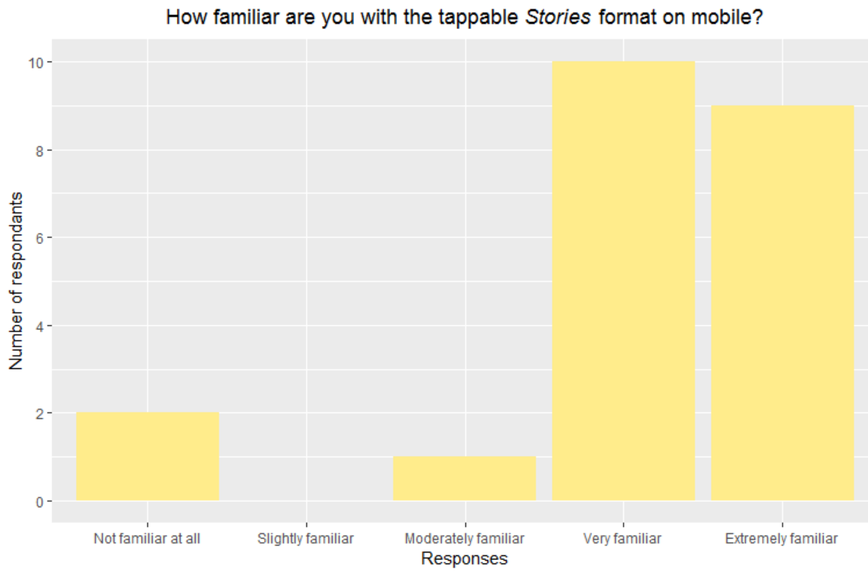


Figure 16: Participant breakdown by familiarity with Stories.

User feedback regarding why Video kept their attention better
i liked the sound
voice was engaging, my attention was directed more automatically to where it should be
Easier to pay attention to the video and listen to the commentary instead of the stories where I had to direct attention to words and then to the video in the background
It is easier to watch a video and listen to a voice than read
The sound mostly. I also found his voice entertaining because it wasn't monotone. What I do like about the stories though is the ability to swipe up in the future and get more information about certain aspects of the snap that someone finds interesting. I also find it easier using the snap version to return to where I was and pick back up where as with the video I'm less likely to do that.

Table 10: Feedback from users who stated that Video kept their attention better than Stories.

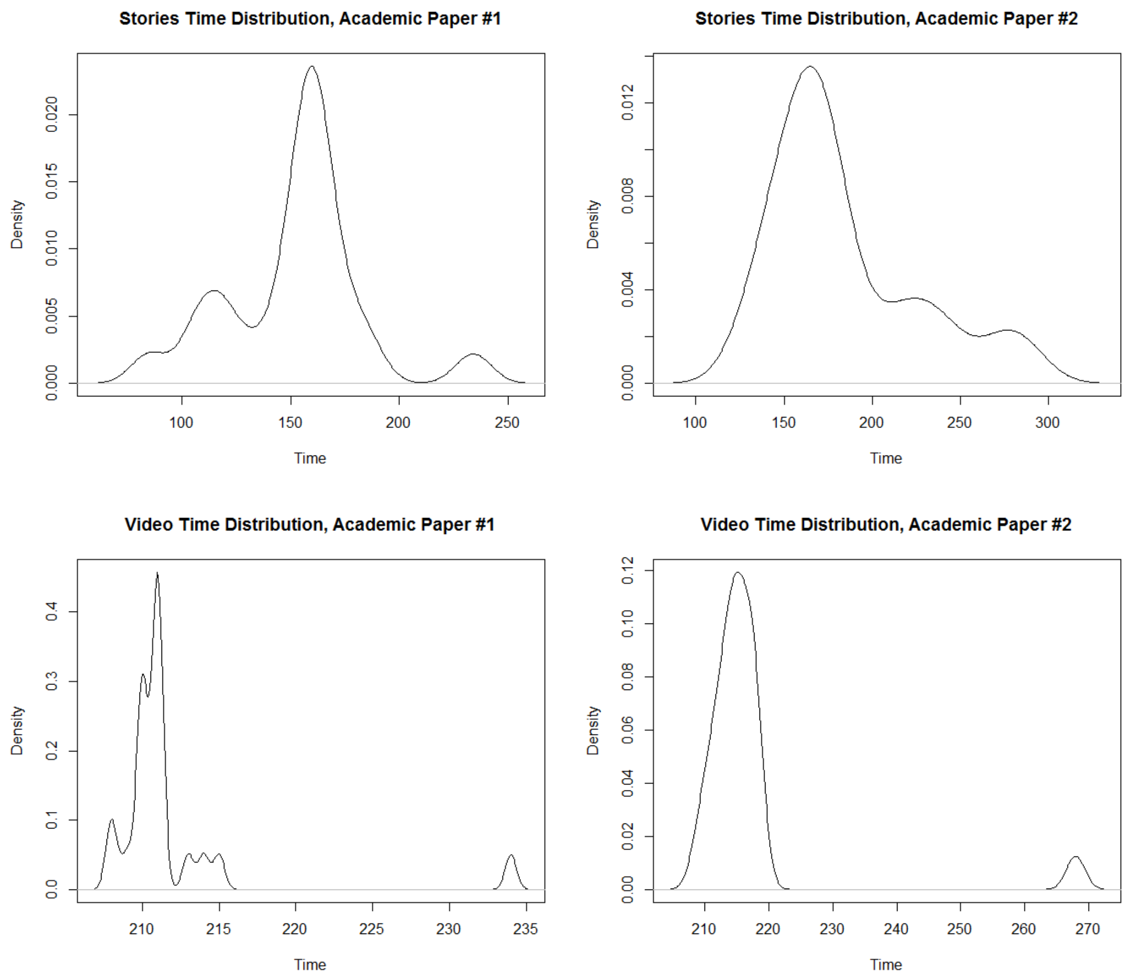


Figure 17: Time distributions of Stories and Video as kernel density plots.

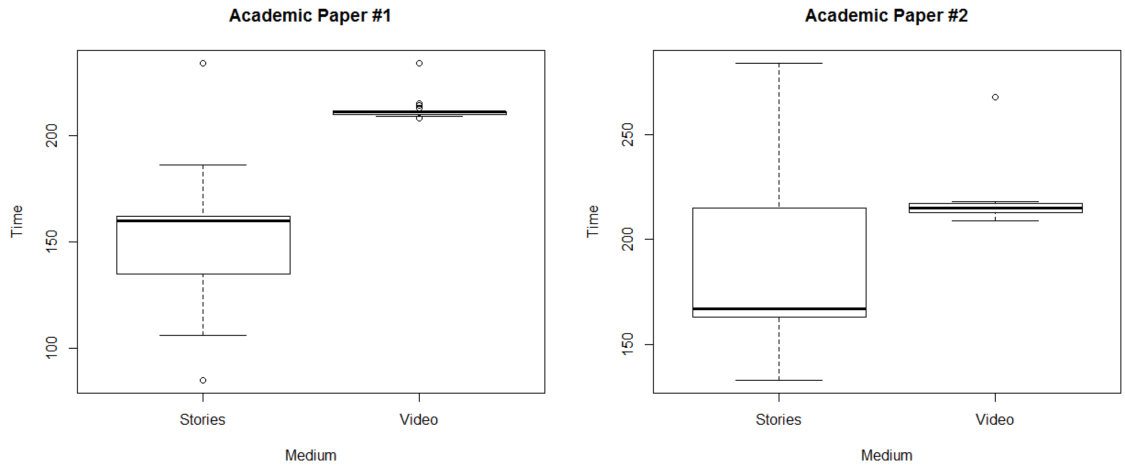


Figure 18: Boxplots of the distribution of time taken to consume content for Stories and Video.

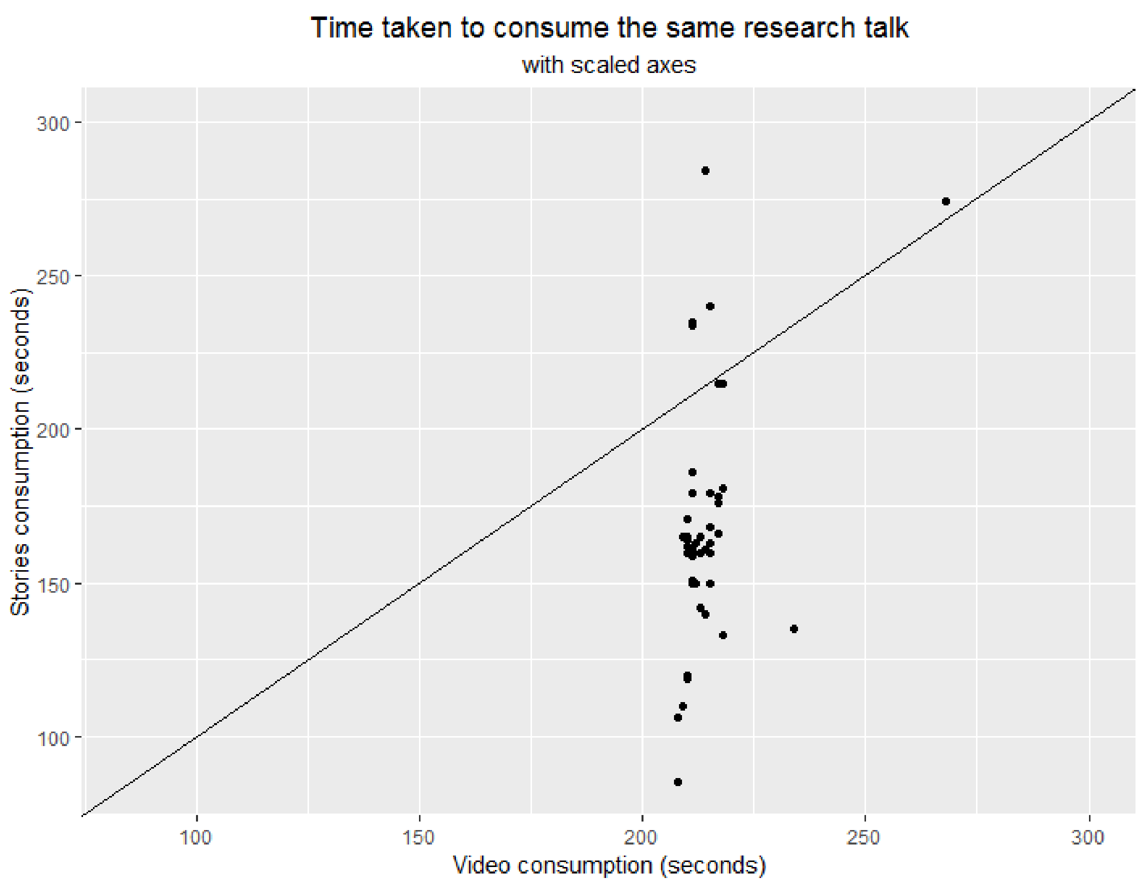


Figure 19: Scatterplot with scaled axes of time taken to consume content by medium plotted among the line y = x.

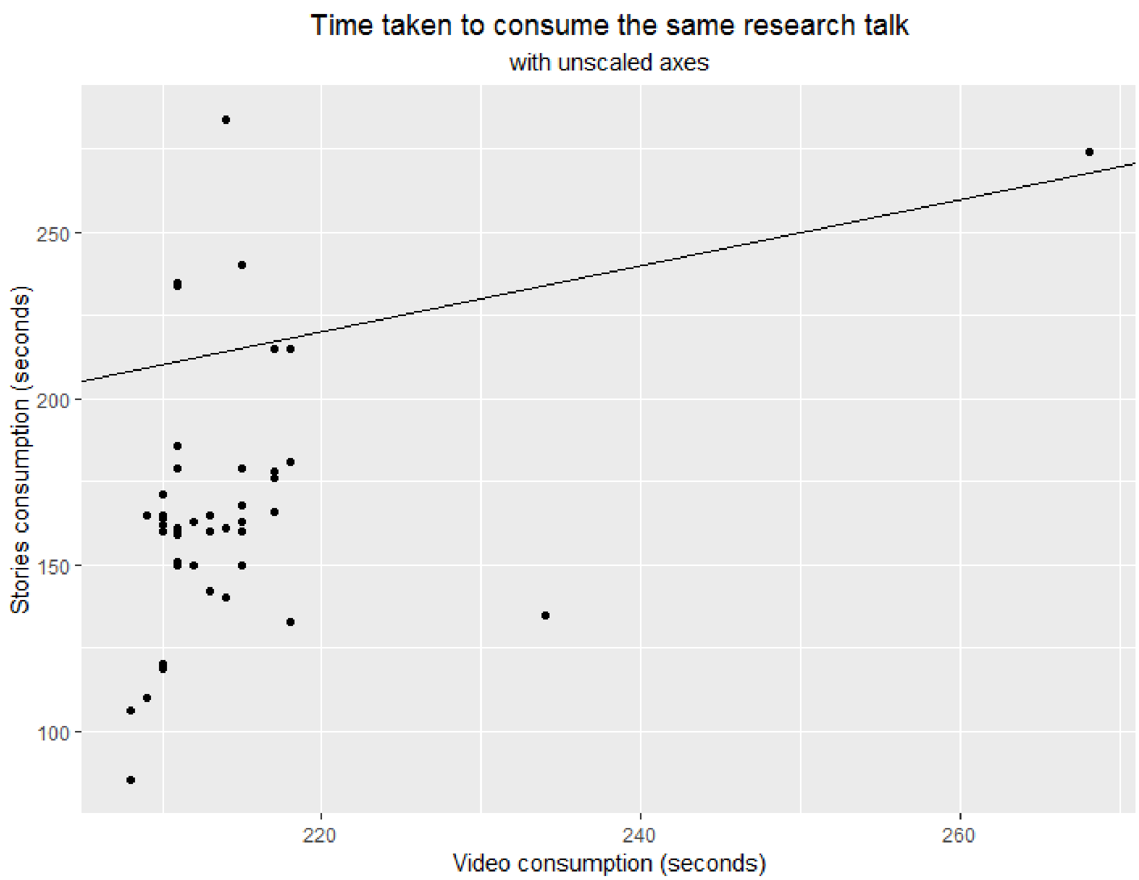


Figure 20: Scatterplot with unscaled axes of time taken to consume content by medium plotted among the line y = x.

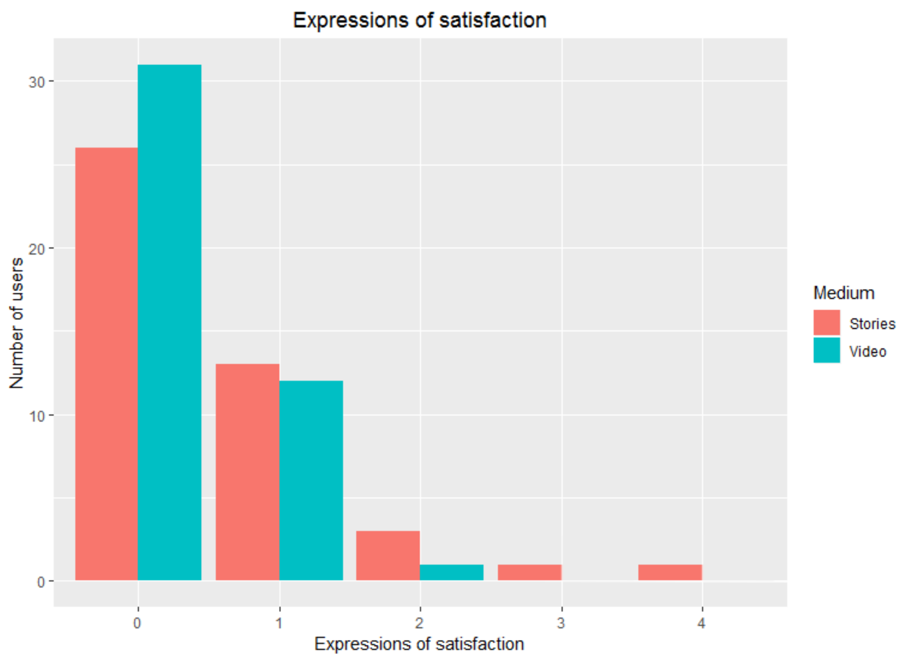


Figure 21: Distribution of expressions of satisfaction by medium.

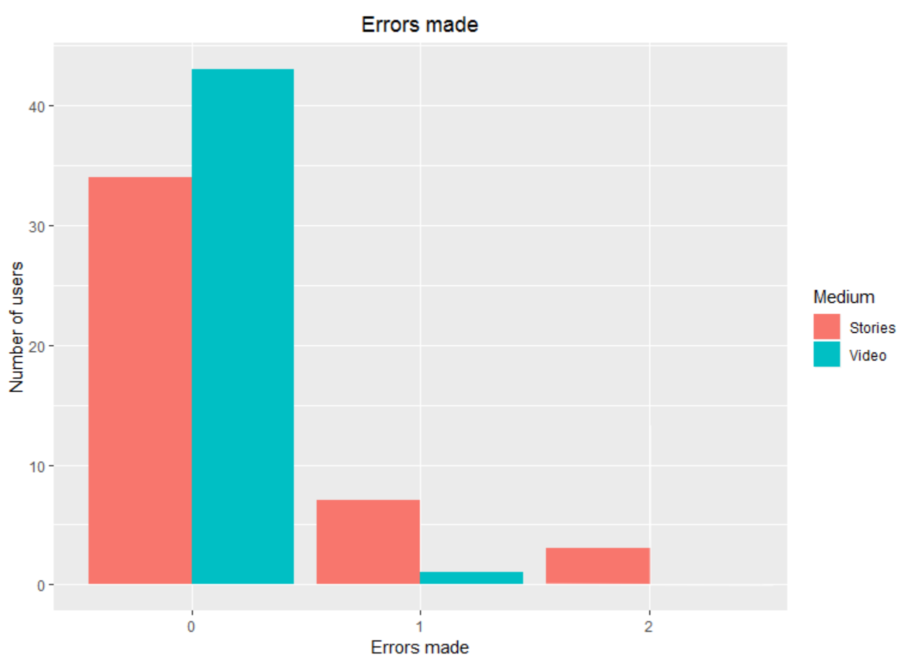


Figure 22: Distribution of errors made by medium.

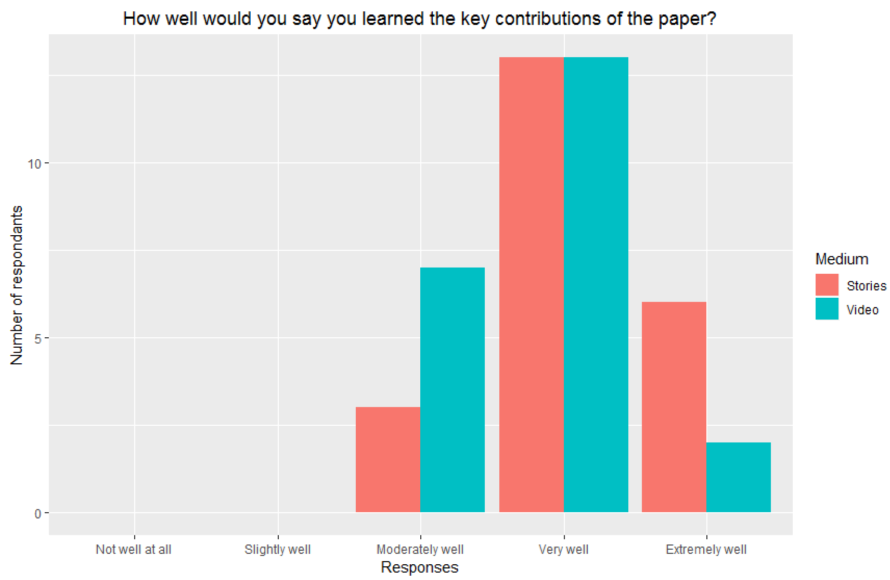


Figure 23: Responses for understanding of key contributions by medium.

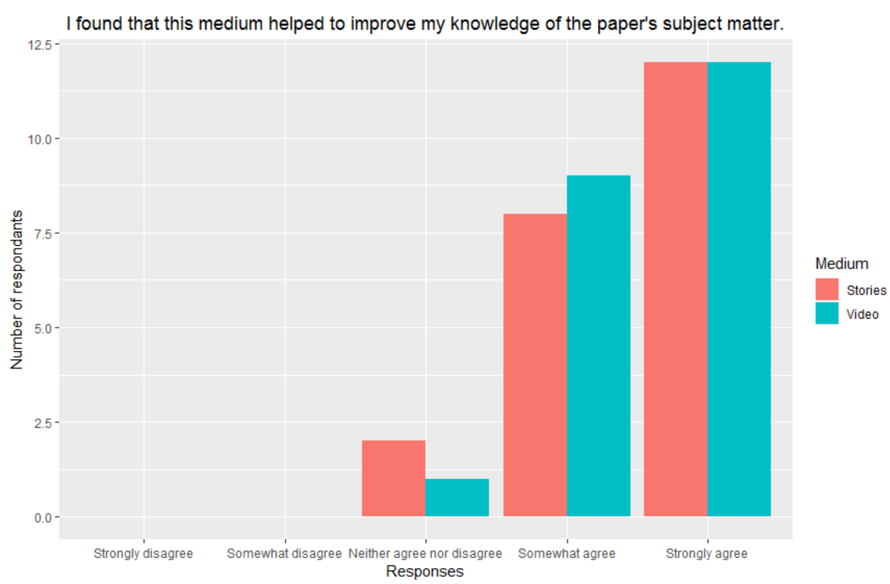


Figure 24: Responses for improvement in knowledge of paper's subject matter by medium.

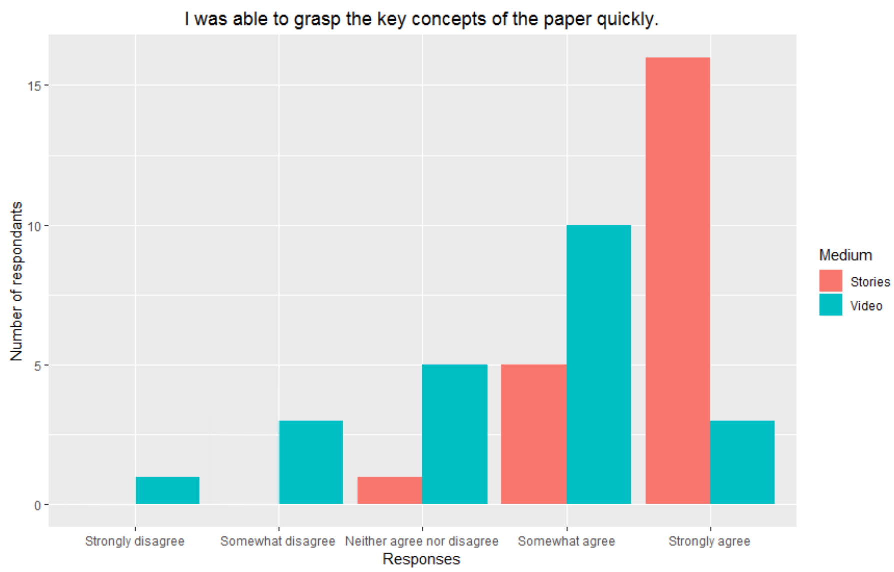


Figure 25: Responses for pace of comprehension by medium.

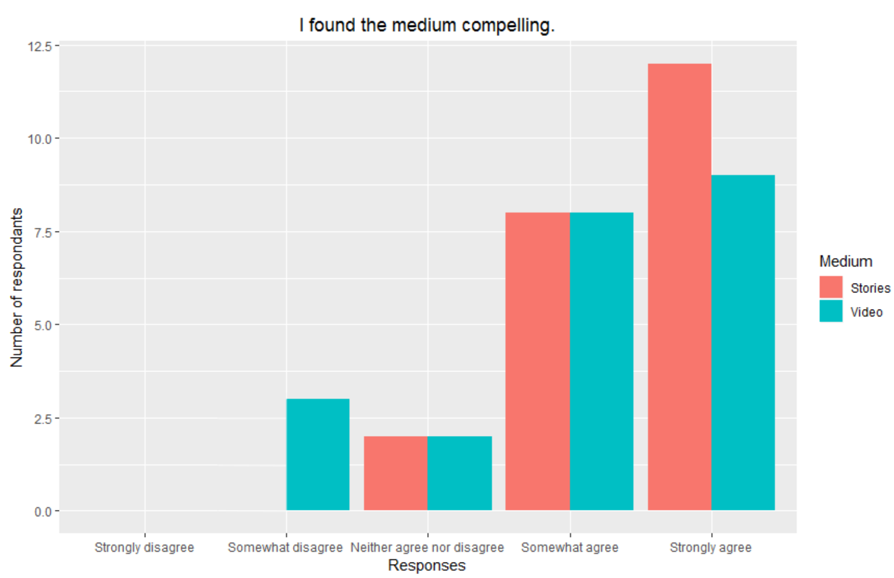


Figure 26: Responses for compelling rating by medium.

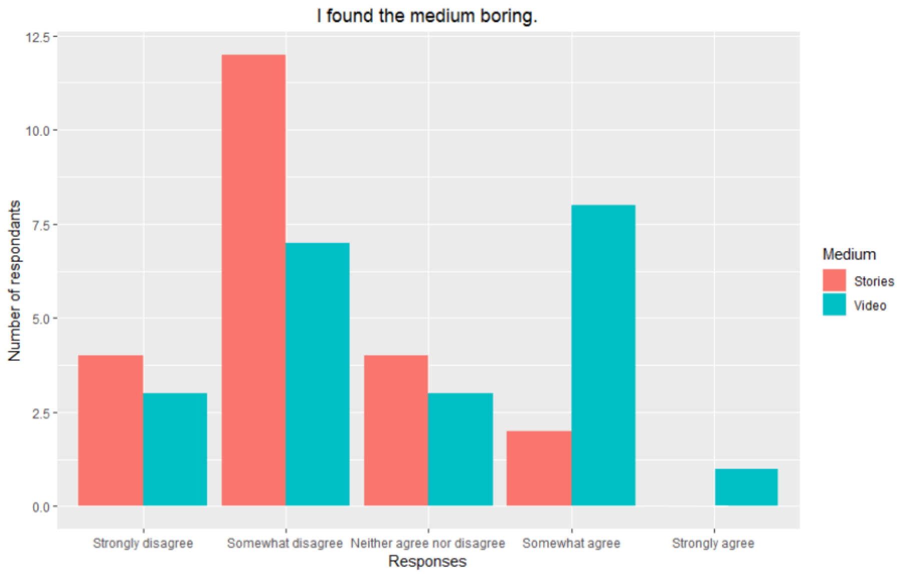


Figure 27: Responses for boring rating by medium.

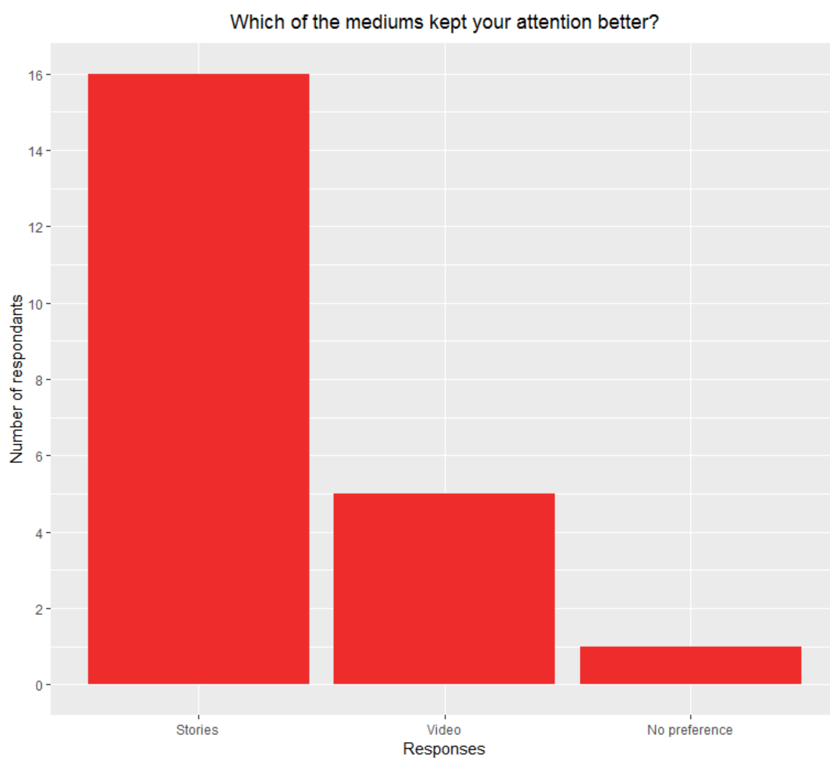


Figure 28: User responses for which format kept their attention better.

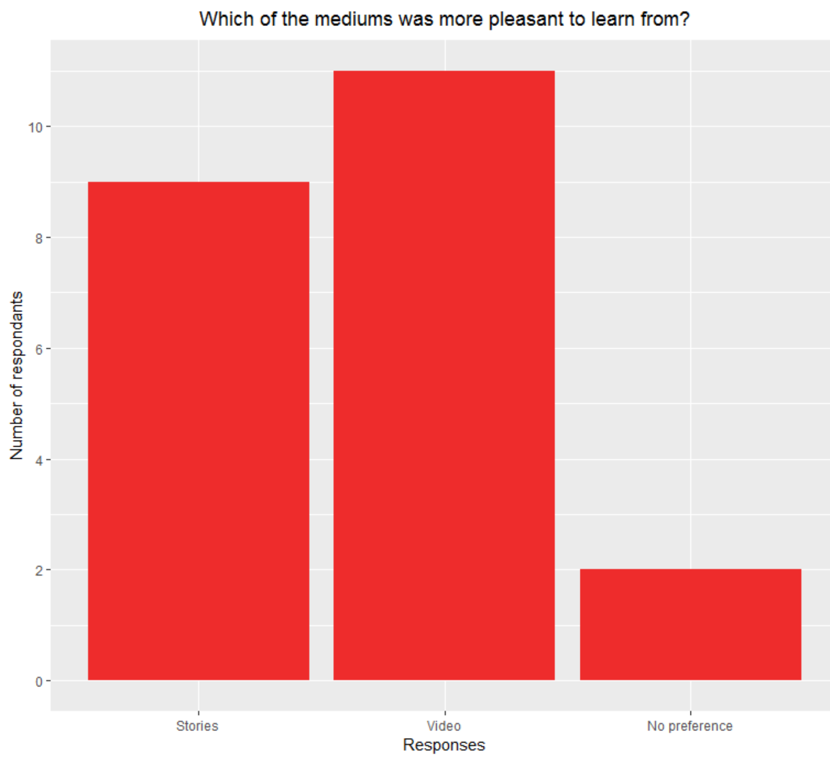


Figure 29: User responses for which format they found more pleasant to learn from.

User feedback for why Stories were more pleasant to learn from
I'm a slow absorber of information so the stories were better to go at my own pace. I would have liked the stories to include some audio.
They are easy to navigate and condensed the information step by step
Easier to go at own pace; more easily digestible than a single constant stream
The story was easy to understand, interactive and I could go at my own pace. The voice in the video was boring, hard to follow and didn't allow me to grasp what was happening due to the pace he was going at.
While there wasn't a voice speaking to me, with tone intonations which are pleasant to listen to, in the stories, I thought the stories were more pleasant because of how they displayed both text and images in a structured way that was pleasant to follow visually.
Despite lacking the in-frame experiment footage provided in the video version, the stories were more pleasant to learn from because I did not have to listen and read at the same time. This allowed for stress-free learning and increased retention.
Stories allowed me to go at my own pace.
Probably the stories because they were more engaging and gave just the key take aways. However, at times the text/description took a while to load which was frustrating because it took longer.
I had more control, feel like I could go about learning at my own pace, my own way.

Table 11: Feedback from users who stated that Stories were more pleasant to learn from than Video.

User feedback for why Video was more pleasant to learn from
Having a narrator made me feel like I was being taught. Hearing the positive tone in the narrator's voice made it pleasant and exciting
It was easier to learn from the video since you do not have to read; you are essentially being taught.
sound
Even though I like to go at my own pace while learning new information, I often like to have audio and visual components when learning new material.
Less effort involved in the overall experience, which led to a pleasant experience.
I liked the video more because in this study but I think it is primarily because the content was relation to vision work and thus having videos of the demo's were really helpful. For domain transfer / adaptation it helps to have two examples next to each other and that is more difficult to do with the snap version. (Nature of content is what led to video preference).
I generally prefer listening to reading, unless the content is particularly challenging
The script from the narrator helped me feel like I understood the topic better, he gave a little extra context to everything
I like having audio and visuals
Having a narrator to explain things added more value to the visuals and it added character.

Table 12: Feedback from users who stated that Video was more pleasant to learn from than Video.

User feedback for having no preference for learning from a particular medium
I like both for different occasions. If I want to effectively focus then stories are preferential and therefore more pleasant. But if I don't want to physically engage and simply get the gist (for review or relaxed learning), then I prefer video.
I thought that both formats were equally pleasant to learn from. Both the video and story formats both contained the same gifs and images that made the information enjoyable.

Table 13: Feedback from users who expressed no preference regarding which medium was more pleasant to learn from.

References

[1] CS147: Intro to Human-Computer Interaction Design. Online; accessed 2019-2-24.

[2] Papers We Love. https://paperswelove.org, 2019. Online; accessed 2018-10-16.

[3] Snapchat Ads. https://forbusiness.snapchat.com/home, 2019. Online; accessed 2019-4-20.

[4] AdEspresso. How the Best Brands Are Crafting Their Instagram Stories. https://adespresso.com/blog/instagram-stories-best-brands-examples-to-copy/, 2018. Online; accessed 2019-4-23.

[5] AMP Open Source Project. AMP. https://amp.dev/, 2019. Online; accessed 2019-4-23.

[6] AMP Open Source Project. AMP Stories. https://amp.dev/about/stories, 2019. Online; accessed 2019-4-23.

[7] Lai Hung Auyeung. Building a collaborative online learning community: A case study in hong kong. *Journal of Educational Computing Research*, 31(2):119–136, 2004.

[8] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 313–322. ACM, 2010.

[9] Peter A. Bruck, Luvai Motiwalla, and Florian Foerster. Mobile learning with micro-content: A framework and evaluation. *Bled eConference*, 25:527–543, 2012.

[10] Miranda G. Capra. Contemporaneous versus retrospective user-reported critical incidents in usability evaluation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(24):1973–1977, 2002.

[11] Michael G. Christel, Michael A. Smith, C. Roy Taylor, and David B. Winkler. Evolving video skims into useful multimedia abstractions. 1998.

[12] Steven M. Drucker, Asta Glatzer, Steven De Mar, and Curtis Wong. Smartskip: consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226. ACM, 2002.

[13] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice.* Springer, Dordrecht, 2007.

[14] Joseph S. Dumas and Janice C. Redish. *A Practical Guide to Usability Testing.* Intellect Books, Exeter, UK, 1st edition, 1999.

[15] Darrell Etherington. Snapchat Gets Its Own Timeline With Snapchat Stories, 24-Hour Photo & Video Tales. https://techcrunch.com, 2013. Online; accessed 2018-4-20.

[16] Facebook IQ. Capturing attention in feed: The science behind effective video creative, April 2016.

[17] Facebook IQ. Why creativity matters more in the age of mobile, February 2017.

[18] Rudy Galfi. AMP stories: Bringing visual storytelling to the open web. https://developers.googleblog.com, 2018. Online; accessed 2018-4-20.

[19] Gene V. Glass, Percy D. Peckham, and James R. Sanders. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, 42(3):237–288, 1972.

[20] Sharad Goel, Amin Saberi, Rajan Vaish, and Shirish Goyal. Stanford Scholar. https://scholar.stanford.edu, 2017. Online; accessed 2018-10-8.

[21] Anuradha A. Gokhale. Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7(1), 1995.

[22] John Green and Hank Green. Crash Course. https://www.youtube.com/user/crashcourse, 2019. Online; accessed 2018-10-16.

[23] Ellis Hamburger. Surprise: Snapchat's most popular feature isn't snaps anymore. https://www.theverge.com, 2014. Online; accessed 2018-4-20.

[24] Lynnsay Hickman, Allen R. Firestone, F. Michael Beck, and Shari Speer. Eye fixations when viewing faces. *The Journal of the American Dental Association*, 141(1):40–46, 2010.

[25] Andreas Holzinger. Usability engineering methods for software developers. *Commun. ACM*, 48(1):71–74, January 2005.

[26] Susan Jamieson et al. Likert scales: how to (ab) use them. *Medical education*, 38(12):1217–1218, 2004.

[27] S. Kennedy. Using video in the bnr usability lab. *SIGCHI Bull.*, 21(2):92–95, October 1989.

[28] Gideon Keren. Between- or within-subjects design: A methodological dilemma. *A Handbook for Data Analysis in the Behaviorial Sciences*, 1:257–272, 2014.

[29] Salman Khan. Khan Academy. https://www.youtube.com/khanacademy, 2019. Online; accessed 2018-10-16.

[30] Herbert E. Krugman and Eugene L. Hartley. Passive learning from television. *Public Opinion Quarterly*, 34(2):184–190, 1970.

[31] Mike Kuniavsky, Elizabeth Goodman, and Andrea Moed. *Observing the User Experience: A Practitioner's Guide to User Research*. Elsevier Science & Technology, Saint Louis, 2012.

[32] Diane Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.

[33] Gitta H. Lubke and Bengt O. Muthén. Applying multigroup confirmatory factor models for continuous outcomes to likert scale data complicates meaningful group comparisons. *Structural equation modeling*, 11(4):514–534, 2004.

[34] Jeffrey Nevid. Teaching the millennials. *APS Observer*, 24:53–56, 2011.

[35] Jon Newmuis and Wei Hong. AMP Stories: The Story so far (AMP Conf '19).

[36] Jon Newmuis and Elite Truong. Telling Stories with AMP.

[37] Jakob Nielsen. *Usability engineering*. Elsevier, 1994.

[38] Jakob Nielsen and Robert L. Mack. *Usability Inspection Methods*. Wiley, 1994.

[39] Jakob Nielsen and Kara Pernice. *Eyetracking Web Usability*. New Riders, 2009.

[40] Stanley Le Baron Payne. *The Art of Asking Questions: Studies in Public Opinion, 3*, volume 451. Princeton University Press, 2014.

[41] Pearson. Beyond millennials: The next generation of learners. https://www.pearsonlearned.com/beyond-millennials-the-next-generation-of-learners, 2018. Online; accessed 2018-12-5.

[42] Pete Prestipino. Create Visual Stories with Google's New AMP Format, 2018. Online; accessed 2019-4-23.

[43] Shalini Ramachandran. Media companies line up to make shows for snap tv. https://www.wsj.com, 2017. Online; accessed 2018-4-20.

[44] Mike Sharples, Josie Taylor, and Giasemi Vavoula. A theory of learning for the mobile age. In *Medienbildung in neuen Kulturräumen*, pages 87–99. Springer, 2010.

[45] Statista. Most popular social media sites used by U.S. teens and young adults 2017. https://www.statista.com/statistics/199242/social-media-and-networking-sites-used-by-us-teenagers/, 2017. Online; accessed 2019-4-23.

[46] Statista. Preferred social networks of U.S. teens 2012-2018. https://www.statista.com/statistics/250172/social-network-usage-of-us-teens-and-young-adults/, 2018. Online; accessed 2019-4-23.

[47] S. Sudman and N.M. Bradburn. *Asking questions: A practical guide to questionnaire design.* Jossey-Bass, 1982.

[48] Jaime Teevan, Shamsi T. Iqbal, Carrie J. Cai, Jeffrey P. Bigham, Michael S. Bernstein, and Elizabeth M. Gerber. Productivity decomposed: Getting big things done with little microtasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 3500–3507, New York, NY, USA, 2016. ACM.

[49] Toney Tse, Sandor Vegh, Ben Shneiderman, and Gary Marchionini. An exploratory study of video browsing, user interface designs and research methodologies: Effectiveness in information seeking tasks. In *Proceedings of the Annual Meeting-American Society for Information Science*, volume 36, pages 681–692. Information Today; 1998, 1999.

[50] Rajan Vaish, Shirish Goyal, Amin Saberi, and Sharad Goel. Creating crowdsourced research talks at scale. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 1–11. International World Wide Web Conferences Steering Committee, 2018.

[51] Jacob O. Wobbrock. Practical Statistics for Human-Computer Interaction. Online; accessed 2019-3-10.

[52] Karoly Zsolnai-Feher. Two Minute Papers. https://www.youtube.com/user/keeroyz, 2019. Online; accessed 2018-10-16.

A R Code

```r
# Ieva Burk
# Senior Thesis
# Scenario: Consuming academic research through two different
# mobile mediums, Stories and Video
```

```r
# DATA SET 1: Performance measures
```

```r
# Read in a data file
studyrecordings = read.csv("studyrecordings.csv")
# Filter by academic paper
studyrecordings_ai <- studyrecordings %>%
                            filter(Academic_paper == "AI")
studyrecordings_cv <- studyrecordings %>%
                            filter(Academic_paper == "CV")
```

```r
# Academic paper #1
# Convert User and Order to a categorical factor
studyrecordings_ai$User = factor(studyrecordings_ai$User)
studyrecordings_ai$Order = factor(studyrecordings_ai$Order)
summary(studyrecordings_ai)
```

```
##       User    Academic_paper Order      Medium        Time
##   1      : 2   AI:44          1:32    Stories:22   Min.   : 85.0
##   2      : 2   CV: 0          2:12    Video  :22   1st Qu.:160.0
##   3      : 2                                       Median :208.0
##   4      : 2                                       Mean   :181.9
##   5      : 2                                       3rd Qu.:211.0
##   6      : 2                                       Max.   :234.0
##   (Other):32
##    Errors_made     Expressions_of_satisfaction
##   Min.   :0.0000   Min.   :0.0000
##   1st Qu.:0.0000   1st Qu.:0.0000
##   Median :0.0000   Median :0.0000
##   Mean   :0.1364   Mean   :0.5227
##   3rd Qu.:0.0000   3rd Qu.:1.0000
##   Max.   :1.0000   Max.   :4.0000
##
```

```r
# Academic paper #2
studyrecordings_cv$User = factor(studyrecordings_cv$User)
```

```r
studyrecordings_cv$Order = factor(studyrecordings_cv$Order)
summary(studyrecordings_cv)
```

```
##       User    Academic_paper Order      Medium         Time      
##  1       : 2   AI: 0          1:32   Stories:22   Min.   :133.0  
##  2       : 2   CV:44          2:12   Video  :22   1st Qu.:167.5  
##  3       : 2                                      Median :213.0  
##  4       : 2                                      Mean   :200.4  
##  5       : 2                                      3rd Qu.:217.0  
##  6       : 2                                      Max.   :284.0  
##  (Other):32                                                      
##   Errors_made     Expressions_of_satisfaction
##  Min.   :0.0000   Min.   :0.0000             
##  1st Qu.:0.0000   1st Qu.:0.0000             
##  Median :0.0000   Median :0.0000             
##  Mean   :0.1818   Mean   :0.3864             
##  3rd Qu.:0.0000   3rd Qu.:1.0000             
##  Max.   :2.0000   Max.   :3.0000             
## 
```

```r
# General
studyrecordings$User = factor(studyrecordings$User)
studyrecordings$Order = factor(studyrecordings$Order)
summary(studyrecordings)
```

```
##       User    Academic_paper Order      Medium         Time      
##  1       : 4   AI:44          1:64   Stories:44   Min.   : 85.0  
##  2       : 4   CV:44          2:24   Video  :44   1st Qu.:162.8  
##  3       : 4                                      Median :210.0  
##  4       : 4                                      Mean   :191.2  
##  5       : 4                                      3rd Qu.:214.2  
##  6       : 4                                      Max.   :284.0  
##  (Other):64                                                      
##   Errors_made     Expressions_of_satisfaction
##  Min.   :0.0000   Min.   :0.0000             
##  1st Qu.:0.0000   1st Qu.:0.0000             
##  Median :0.0000   Median :0.0000             
##  Mean   :0.1591   Mean   :0.4545             
##  3rd Qu.:0.0000   3rd Qu.:1.0000             
##  Max.   :2.0000   Max.   :4.0000             
## 
```

```r
# Numerical variable analysis
# VARIABLE: Time to consume content
# Descriptive statistics by Medium
library(plyr)
ddply(studyrecordings_ai, ~ Medium, function(data) summary(data$Time))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories   85  138.75    160 152.0000  161.75  234
## 2   Video  208  210.00    211 211.8182  211.00  234
```

```r
ddply(studyrecordings_ai, ~ Medium, summarize, Time.mean=mean(Time),
Time.sd=sd(Time))
```

```
##    Medium Time.mean   Time.sd
## 1 Stories  152.0000 31.233377
## 2   Video  211.8182  5.215694
```

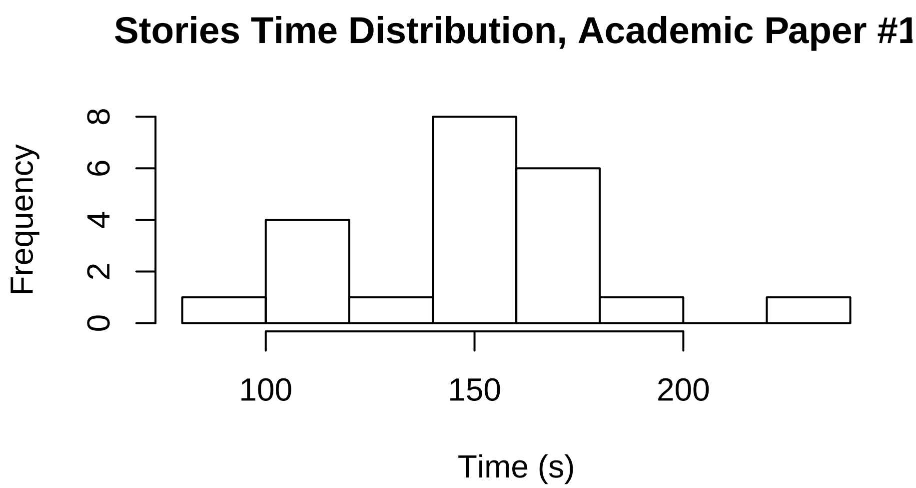
```r
ddply(studyrecordings_cv, ~ Medium, function(data) summary(data$Time))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories  133     163    167 183.9091   206.5  284
## 2   Video  209     213    215 216.9545   217.0  268
```

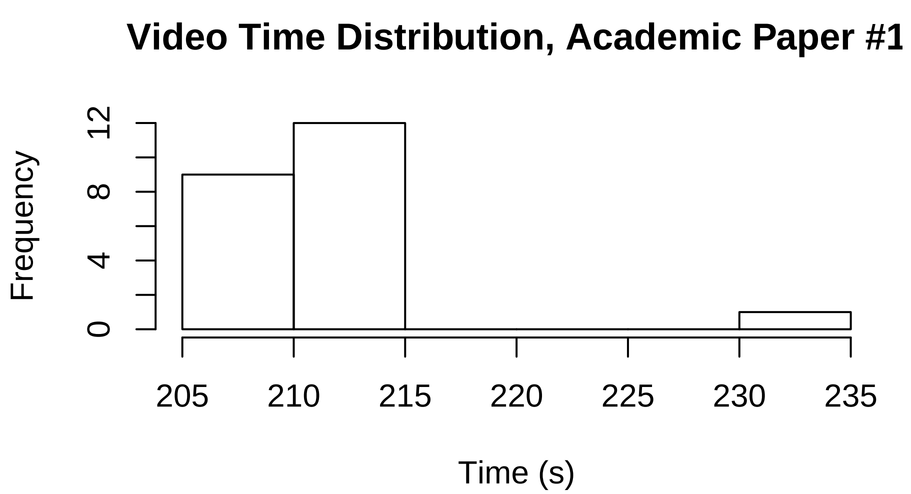
```r
ddply(studyrecordings_cv, ~ Medium, summarize, Time.mean=mean(Time),
Time.sd=sd(Time))
```

```
##    Medium Time.mean  Time.sd
## 1 Stories  183.9091 42.02597
## 2   Video  216.9545 11.69647
```

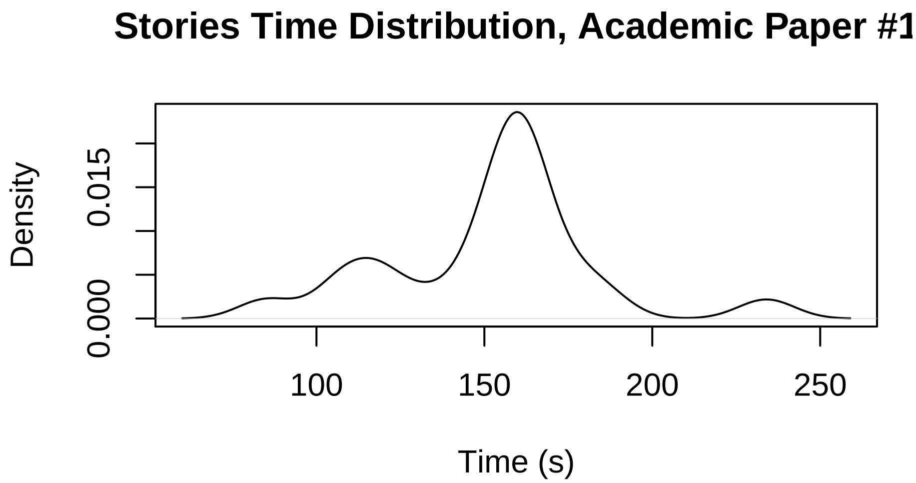
```r
# Data distributions and boxplots for the Time responses
# Academic paper #1
# Histogram
hist(studyrecordings_ai[studyrecordings_ai$Medium == "Stories",]$Time,
     main = "Stories Time Distribution, Academic Paper #1",
     xlab = "Time (s)")
```



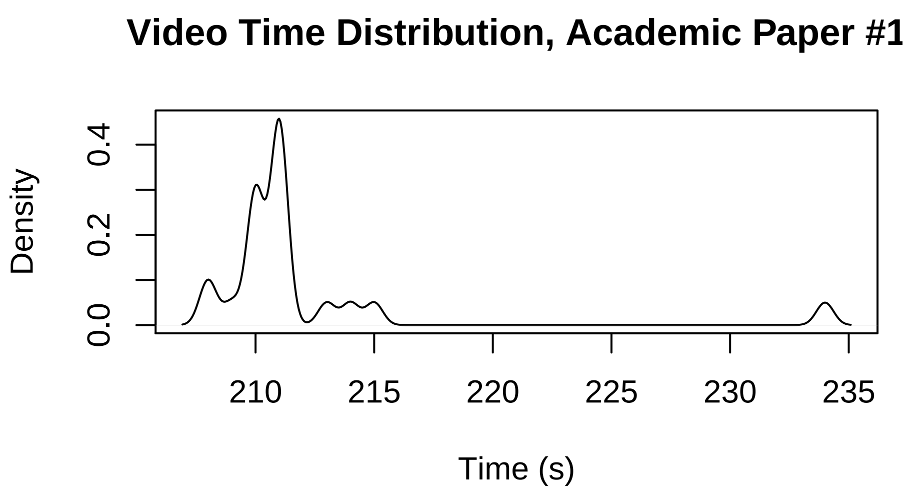

```r
hist(studyrecordings_ai[studyrecordings_ai$Medium == "Video",]$Time,
     main = "Video Time Distribution, Academic Paper #1",
     xlab = "Time (s)")
```



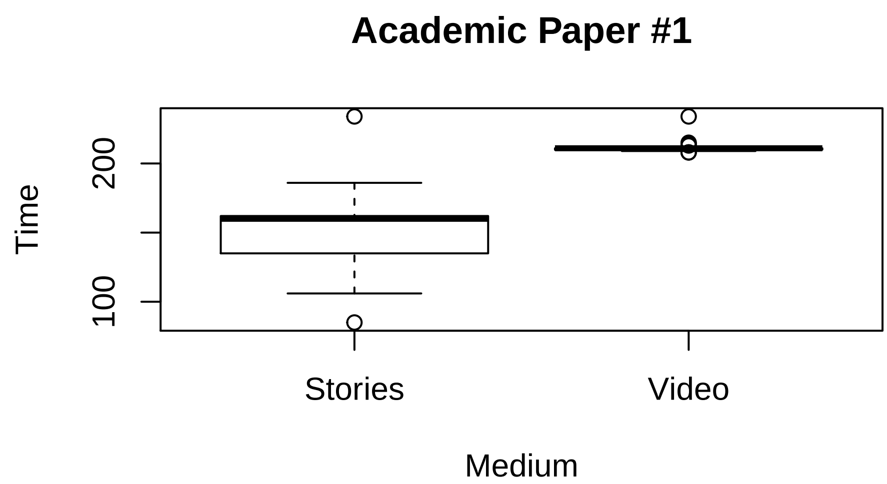

```r
# Kernel density plot
plot(density
(studyrecordings_ai[studyrecordings_ai$Medium == "Stories",]$Time),
     main = "Stories Time Distribution, Academic Paper #1",
     xlab = "Time (s)")
```



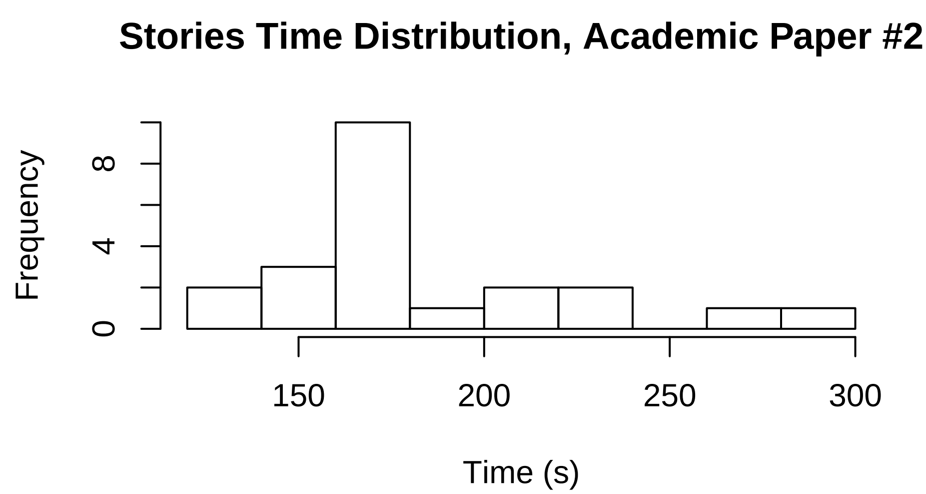

```
plot(density
    (studyrecordings_ai[studyrecordings_ai$Medium == "Video",]$Time),
     main = "Video Time Distribution, Academic Paper #1",
     xlab = "Time (s)")
```



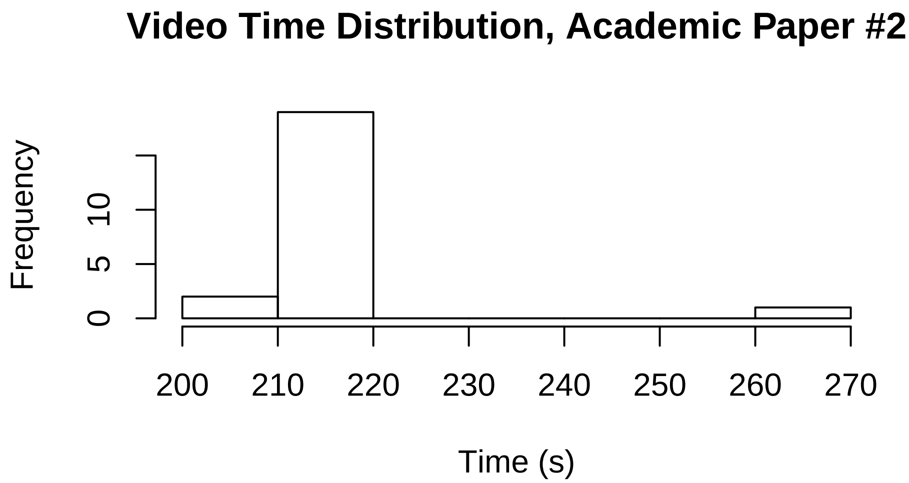

```
# Boxplot
plot(Time ~ Medium, data=studyrecordings_ai,
     main = "Academic Paper #1")
```



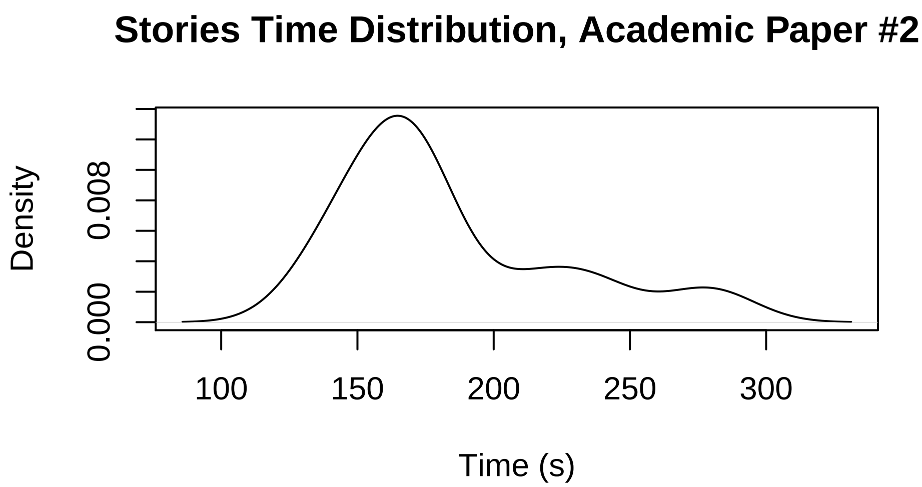

```r
# Academic paper #2
# Histogram
hist(studyrecordings_cv[studyrecordings_cv$Medium == "Stories",]$Time,
     main = "Stories Time Distribution, Academic Paper #2",
     xlab = "Time (s)")
```



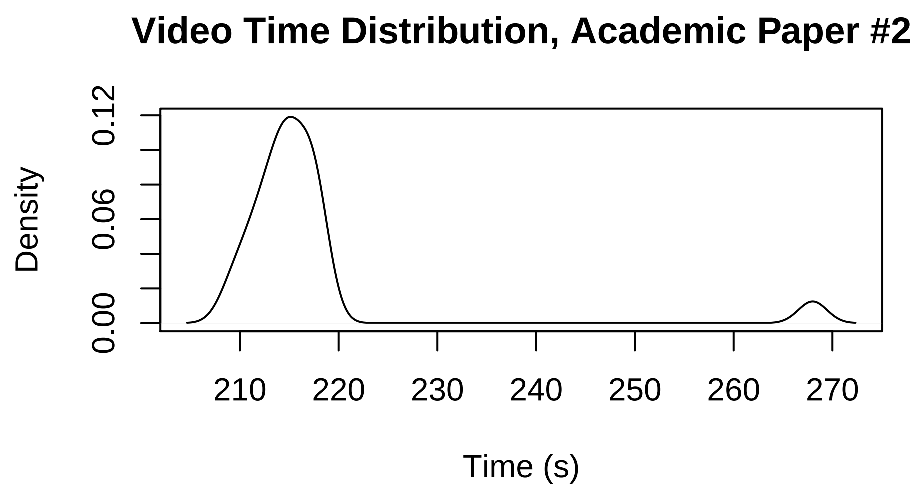

```r
hist(studyrecordings_cv[studyrecordings_cv$Medium == "Video",]$Time,
     main = "Video Time Distribution, Academic Paper #2",
     xlab = "Time (s)")
```



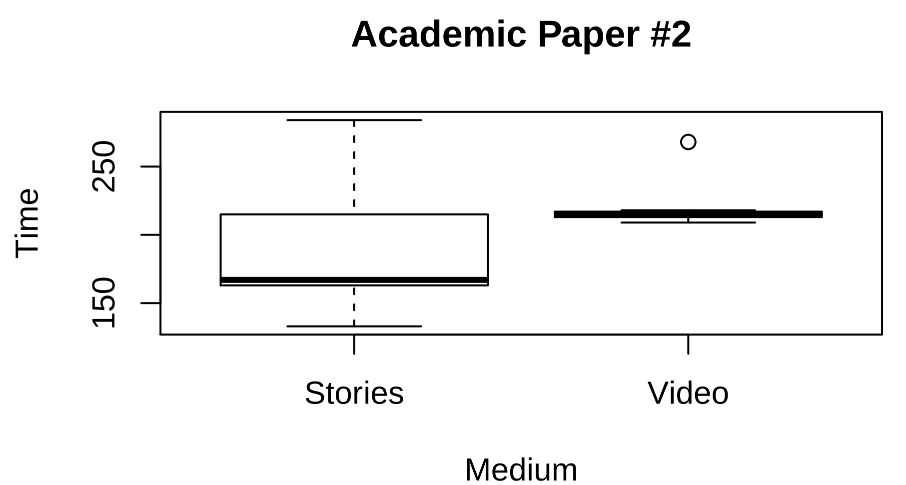

```r
# Kernel density plot
plot(density
(studyrecordings_cv[studyrecordings_cv$Medium == "Stories",]$Time),
     main = "Stories Time Distribution, Academic Paper #2",
     xlab = "Time (s)")
```



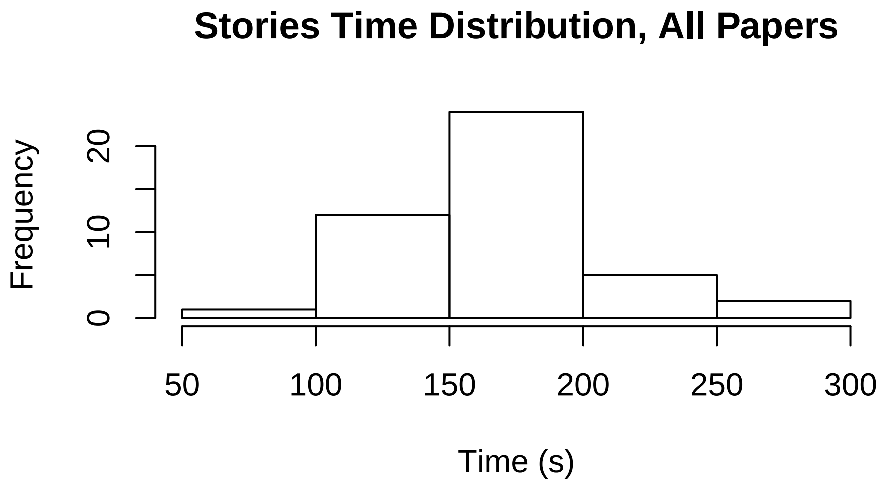

```r
plot(density
(studyrecordings_cv[studyrecordings_cv$Medium == "Video",]$Time),
     main = "Video Time Distribution, Academic Paper #2",
     xlab = "Time (s)")
```



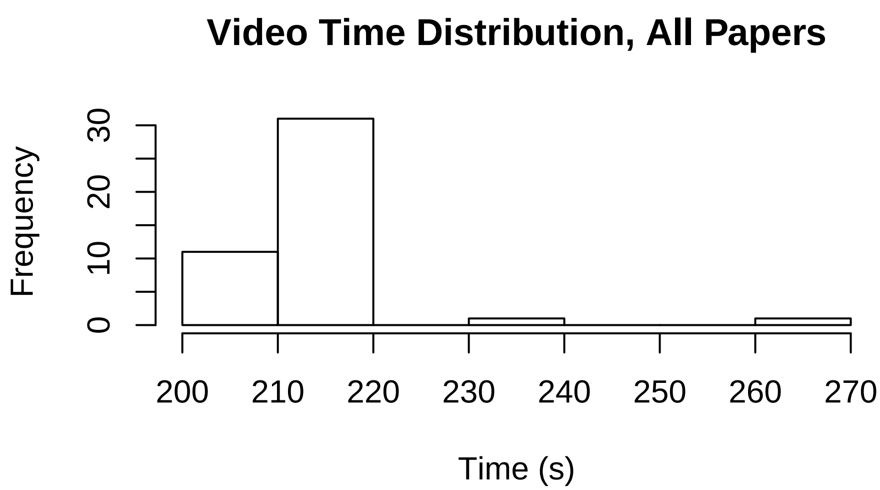

```r
# Boxplot
plot(Time ~ Medium, data=studyrecordings_cv,
     main = "Academic Paper #2")
```



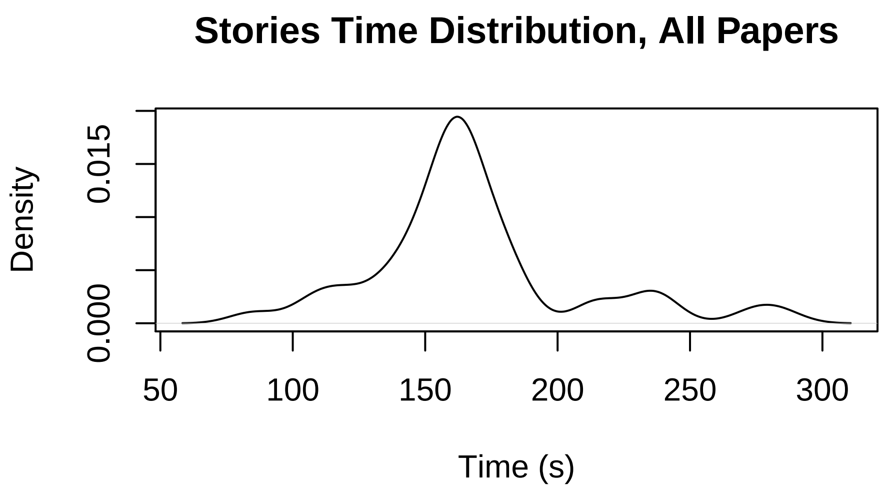

```r
# All papers
# Histogram
hist(studyrecordings[studyrecordings$Medium == "Stories",]$Time,
     main = "Stories Time Distribution, All Papers",
     xlab = "Time (s)") # histogram
```



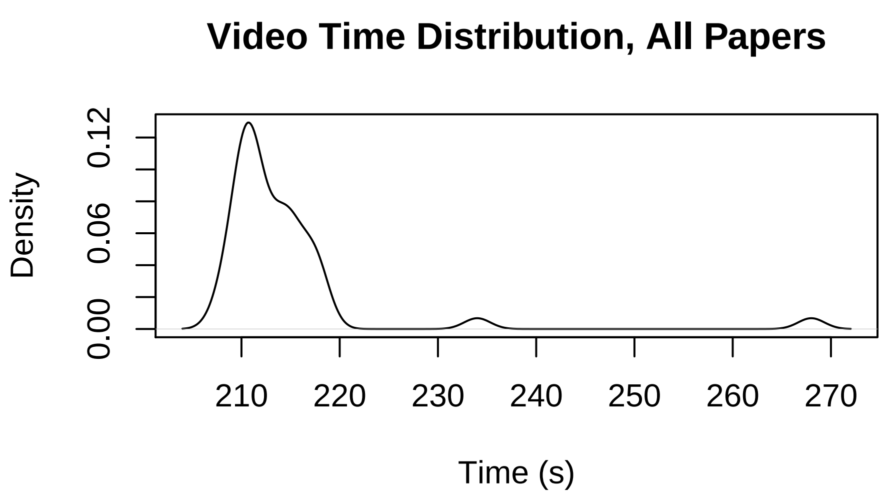

```
hist(studyrecordings[studyrecordings$Medium == "Video",]$Time,
     main = "Video Time Distribution, All Papers",
     xlab = "Time (s)")
```




```
# Kernel density plot
plot(density
    (studyrecordings[studyrecordings$Medium == "Stories",]$Time),
     main = "Stories Time Distribution, All Papers",
     xlab = "Time (s)")
```




```r
plot(density(studyrecordings[studyrecordings$Medium == "Video",]$Time),
     main = "Video Time Distribution, All Papers", xlab = "Time (s)")
```




```r
# Test ANOVA assumptions
# Shapiro-Wilk
# Academic paper #1
shapiro.test(studyrecordings_ai[studyrecordings_ai$Medium ==
 "Stories",]$Time)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  studyrecordings_ai[studyrecordings_ai$Medium == "Stories", ]$Time
## W = 0.91836, p-value = 0.0703
```

```r
shapiro.test(studyrecordings_ai[studyrecordings_ai$Medium ==
 "Video",]$Time)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  studyrecordings_ai[studyrecordings_ai$Medium == "Video", ]$Time
## W = 0.48666, p-value = 9.874e-08
```

```r
# Academic paper #2
shapiro.test(studyrecordings_cv[studyrecordings_cv$Medium ==
 "Stories",]$Time)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  studyrecordings_cv[studyrecordings_cv$Medium == "Stories", ]$Time
## W = 0.8593, p-value = 0.004979
```

```r
shapiro.test(studyrecordings_cv[studyrecordings_cv$Medium ==
 "Video",]$Time)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  studyrecordings_cv[studyrecordings_cv$Medium == "Video", ]$Time
## W = 0.4255, p-value = 2.819e-08
```

```r
# Academic paper #1
# Fitting a model for testing residuals
m = aov(Time ~ Medium + Error(User/Medium), data=studyrecordings_ai)

# Getting residuals for User
shapiro.test(residuals(m$User))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m$User)
## W = 0.89845, p-value = 0.03266
```

```r
# Academic paper #2
m = aov(Time ~ Medium + Error(User/Medium), data=studyrecordings_cv)
shapiro.test(residuals(m$User))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m$User)
## W = 0.852, p-value = 0.0046
```

```r
# Wilcoxon signed-rank test on Time
library(coin)
wilcoxsign_test(Time ~ Medium | User, data=studyrecordings_ai,
distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = -4.0755, p-value = 9.537e-07
## alternative hypothesis: true mu is not equal to 0
```

```r
wilcoxsign_test(Time ~ Medium | User, data=studyrecordings_cv,
distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = -3.1005, p-value = 0.001091
## alternative hypothesis: true mu is not equal to 0
```

```r
# VARIABLE: Expressions of satisfaction
# Descriptive statistics by Medium
library(plyr)
ddply(studyrecordings_ai, ~ Medium, function(data)
summary(data$Expressions_of_satisfaction))
```

```
##    Medium Min. 1st Qu. Median      Mean 3rd Qu. Max.
## 1 Stories    0       0      0 0.6818182       1    4
## 2   Video    0       0      0 0.3636364       1    1
```

```r
ddply(studyrecordings_ai, ~ Medium, summarize,
      Expns_of_satisfaction.mean=mean(Expressions_of_satisfaction),
      Expns_of_satisfaction.sd=sd(Expressions_of_satisfaction))
```

```
##    Medium Expns_of_satisfaction.mean Expns_of_satisfaction.sd
## 1 Stories                  0.6818182                 0.994574
## 2   Video                  0.3636364                 0.492366
```

```r
ddply(studyrecordings_cv, ~ Medium, function(data)
summary(data$Expressions_of_satisfaction))
```

```
##    Medium Min. 1st Qu. Median      Mean 3rd Qu. Max.
## 1 Stories    0       0      0 0.5000000       1    3
## 2   Video    0       0      0 0.2727273       0    2
```

```r
ddply(studyrecordings_cv, ~ Medium, summarize,
      Expns_of_satisfaction.mean=mean(Expressions_of_satisfaction),
      Expns_of_satisfaction.sd=sd(Expressions_of_satisfaction))
```

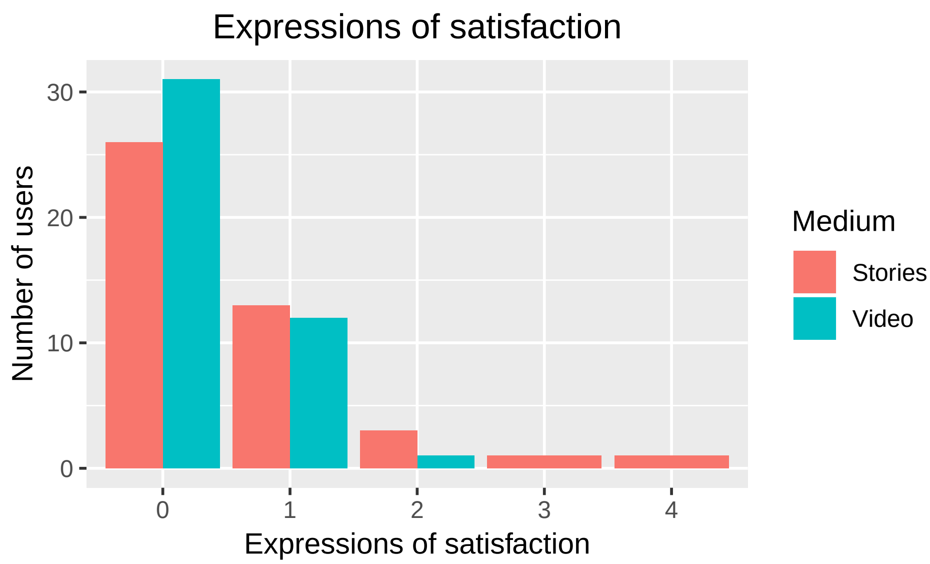
```
##    Medium Expns_of_satisfaction.mean Expns_of_satisfaction.sd
## 1 Stories                  0.5000000                0.8017837
## 2   Video                  0.2727273                0.5504819
```

```r
# Data distributions for Expressions of satisfaction,
# grouped by Medium
# All papers
library(ggplot2)
studyrecordings_leveled <- studyrecordings %>%
  mutate(Expressions_of_satisfaction = factor(Expressions_of_satisfaction,
                                              levels = c("0", "1", "2",
                                                         "3", "4")))
studyrecordings_leveled$Expressions_of_satisfaction
```

```
##  [1] 2 1 0 0 0 0 1 1 2 1 1 0 1 0 0 0 4 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0
## [36] 1 0 0 1 0 0 1 0 0 1 0 0 0 0 1 0 1 0 0 1 3 1 0 0 0 1 2 0 0 0 0 0
## [71] 0 0 1 1 0 0 1 0 0 1 0 0 2 0 0 0 0 0
## Levels: 0 1 2 3 4
```

```r
studyrecordings_leveled %>%
  ggplot() +
  geom_bar(aes(Expressions_of_satisfaction, fill = Medium),
```

```
position = "dodge") +
  xlab("Expressions of satisfaction") +
  ylab("Number of users") +
  ggtitle("Expressions of satisfaction") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
# Wilcoxon signed-rank test on Expressions of satisfaction, by paper
library(coin)
wilcoxsign_test(Expressions_of_satisfaction ~ Medium | User,
data=studyrecordings_ai,
                distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 1.2828, p-value = 0.2485
## alternative hypothesis: true mu is not equal to 0
```

```
wilcoxsign_test(Expressions_of_satisfaction ~ Medium | User,
data=studyrecordings_cv,
                distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
```

```
##   stratified by block
## Z = 0.81812, p-value = 0.4453
## alternative hypothesis: true mu is not equal to 0
```

```
# VARIABLE: Errors made
# Descriptive statistics by Medium
library(plyr)
ddply(studyrecordings_ai, ~ Medium, function(data)
summary(data$Errors_made))
```

```
##    Medium Min. 1st Qu. Median      Mean 3rd Qu. Max.
## 1 Stories    0       0      0 0.2727273    0.75    1
## 2   Video    0       0      0 0.0000000    0.00    0
```

```
ddply(studyrecordings_ai, ~ Medium, summarize,
    Errors_made.mean=mean(Errors_made),
    Errors_made.sd=sd(Errors_made))
```

```
##    Medium Errors_made.mean Errors_made.sd
## 1 Stories        0.2727273      0.4558423
## 2   Video        0.0000000      0.0000000
```

```
ddply(studyrecordings_cv, ~ Medium, function(data)
summary(data$Errors_made))
```

```
##    Medium Min. 1st Qu. Median       Mean 3rd Qu. Max.
## 1 Stories    0       0      0 0.31818182       0    2
## 2   Video    0       0      0 0.04545455       0    1
```

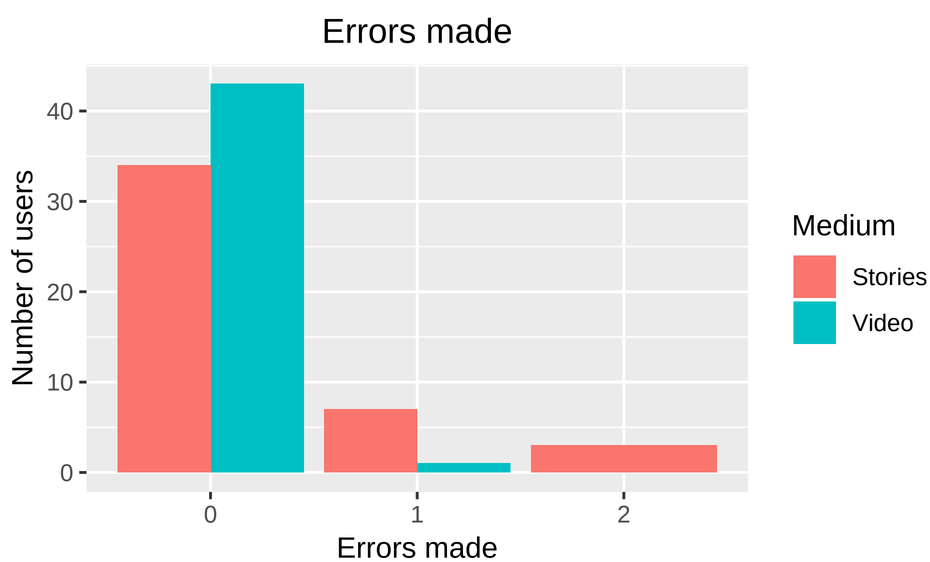
```
ddply(studyrecordings_cv, ~ Medium, summarize,
    Errors_made.mean=mean(Errors_made),
    Errors_made.sd=sd(Errors_made))
```

```
##    Medium Errors_made.mean Errors_made.sd
## 1 Stories       0.31818182      0.7162311
## 2   Video       0.04545455      0.2132007
```

```
# Data distributions for Errors made,
# grouped by medium
# All papers
studyrecordings_leveled <- studyrecordings %>%
  mutate(Errors_made = factor(Errors_made, levels = c("0", "1", "2")))
studyrecordings_leveled$Errors_made
```

```
##  [1] 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
## [36] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 0 0 0 0 2 0 0 1 0 0 0 0 0 0
## [71] 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## Levels: 0 1 2
```

```r
studyrecordings_leveled %>%
  ggplot() +
  geom_bar(aes(Errors_made, fill = Medium), position = "dodge") +
  xlab("Errors made") +
  ylab("Number of users") +
  ggtitle("Errors made") +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
# Wilcoxon signed-rank test on Errors made, by paper
library(coin)
wilcoxsign_test(Errors_made ~ Medium | User, data=studyrecordings_ai,
                distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 2.4495, p-value = 0.03125
## alternative hypothesis: true mu is not equal to 0
```

```r
wilcoxsign_test(Errors_made ~ Medium | User, data=studyrecordings_cv,
                distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 1.9976, p-value = 0.125
## alternative hypothesis: true mu is not equal to 0
```

```r
# Count data only since discrete (only Expressions of satisfaction
# and Errors made). Trying to fit a Poisson distribution for count
# data. Note that ks.test only works for continuous distributions,
# but since Poisson distributions are discrete, use fitdist, not
# fitdistr, and test with gofstat [25].
library(fitdistrplus)
fit = fitdist(studyrecordings
              [studyrecordings$Medium == "Stories",]
              $Expressions_of_satisfaction, "pois",
              discrete=TRUE)
gofstat(fit) # goodness-of-fit test
```

```
## Chi-squared statistic:  0.2555876
## Degree of freedom of the Chi-squared distribution:  1
## Chi-squared p-value:  0.6131679
## Chi-squared table:
##      obscounts theocounts
## <= 0  26.00000   24.36824
## <= 1  13.00000   14.39941
## > 1    5.00000    5.23235
##
## Goodness-of-fit criteria
##                                1-mle-pois
## Akaike's Information Criterion   95.45535
## Bayesian Information Criterion   97.23954
```

```r
fit = fitdist(studyrecordings
              [studyrecordings$Medium == "Video",]
              $Expressions_of_satisfaction, "pois",
              discrete=TRUE)
gofstat(fit) # goodness-of-fit test
```

```
## Chi-squared statistic:  0.7155894
## Degree of freedom of the Chi-squared distribution:  1
## Chi-squared p-value:  0.3975945
##    the p-value may be wrong with some theoretical counts < 5
## Chi-squared table:
##      obscounts theocounts
## <= 0  31.00000   32.00870
## <= 1  12.00000   10.18459
## > 1    1.00000    1.80671
##
## Goodness-of-fit criteria
##                                 1-mle-pois
## Akaike's Information Criterion   63.45000
## Bayesian Information Criterion   65.23419
```

```r
# DATA SET 2: Subjective measures
```

```r
# Read in a data file with all quantitative survey data
library(readr)
survey_quant_raw = read_csv("survey_quant.csv")
```

```r
# Tidy quantitative survey data
library(tidyverse)
require(MASS)
require(dplyr)

survey_quant_tidy <- survey_quant_raw %>%
  dplyr::select(Q3, Q4, Q5, Q6, Q7, Q9, Q11_1, Q11_2, Q12_1, Q12_2,
    Q13_1, Q13_2, Q14_1, Q14_2, Q15_1, Q15_2) %>%
  dplyr::rename("affiliation" = "Q3",
         "area_of_study" = "Q4",
         "stories_familiarity" = "Q5",
         "technical_expertise" = "Q6",
         "attention" = "Q7",
         "learning" = "Q9",
         "contributionunderstanding_Stories" = "Q11_1",
         "contributionunderstanding_Video" = "Q11_2",
         "knowledge_Stories" = "Q12_1",
         "knowledge_Video" = "Q12_2",
         "pace_Stories" = "Q13_1",
         "pace_Video" = "Q13_2",
         "compelling_Stories" = "Q14_1",
         "compelling_Video" = "Q14_2",
```

```r
        "boring_Stories" = "Q15_1",
        "boring_Video" = "Q15_2"
        ) %>%
  mutate(affiliation = recode(affiliation,
                              Other = "Rice Graduate '18")) %>%
  slice(3:n()) %>%
  filter(!is.na(affiliation)) %>%
  mutate(stories_familiarity = factor(stories_familiarity,
                                  levels = c("Not familiar at all",
                                             "Slightly familiar",
                                             "Moderately familiar",
                                             "Very familiar",
                                             "Extremely familiar")),
        contributionunderstanding_Stories = factor
        (contributionunderstanding_Stories,
                                   levels = c("Not well at all",
                                              "Slightly well",
                                              "Moderately well",
                                              "Very well",
                                              "Extremely well")),
        contributionunderstanding_Video = factor
        (contributionunderstanding_Video,
                                     levels = c("Not well at all",
                                                "Slightly well",
                                                "Moderately well",
                                                "Very well",
                                                "Extremely well")),
        knowledge_Stories = factor(knowledge_Stories,
                       levels = c("Strongly disagree",
                                  "Somewhat disagree",
                                  "Neither agree nor disagree",
                                  "Somewhat agree",
                                  "Strongly agree")),
        knowledge_Video = factor(knowledge_Video,
                             levels = c("Strongly disagree",
                                        "Somewhat disagree",
                                        "Neither agree nor disagree",
                                        "Somewhat agree",
                                        "Strongly agree")),
        pace_Stories = factor(pace_Stories,
                              levels = c("Strongly disagree",
                                         "Somewhat disagree",
                                         "Neither agree nor disagree",
```

```
                                         "Somewhat agree",
                                         "Strongly agree")),
         pace_Video = factor(pace_Video,
                            levels = c("Strongly disagree",
                                       "Somewhat disagree",
                                       "Neither agree nor disagree",
                                       "Somewhat agree",
                                       "Strongly agree")),
         compelling_Stories = factor(compelling_Stories,
                            levels = c("Strongly disagree",
                                       "Somewhat disagree",
                                       "Neither agree nor disagree",
                                       "Somewhat agree",
                                       "Strongly agree")),
         compelling_Video = factor(compelling_Video,
                          levels = c("Strongly disagree",
                                       "Somewhat disagree",
                                       "Neither agree nor disagree",
                                       "Somewhat agree",
                                       "Strongly agree")),
         boring_Stories = factor(boring_Stories,
                        levels = c("Strongly disagree",
                                       "Somewhat disagree",
                                       "Neither agree nor disagree",
                                       "Somewhat agree",
                                       "Strongly agree")),
         boring_Video = factor(boring_Video,
                      levels = c("Strongly disagree",
                                       "Somewhat disagree",
                                       "Neither agree nor disagree",
                                       "Somewhat agree",
                                       "Strongly agree"))) %>%
  mutate(stories_familiarity = fct_recode(stories_familiarity,
                                         "1" = "Not familiar at all",
                                         "2" = "Slightly familiar",
                                         "3" = "Moderately familiar",
                                         "4" = "Very familiar",
                                         "5" = "Extremely familiar"),
contributionunderstanding_Stories = fct_recode
(contributionunderstanding_Stories,
                                "1" = "Not well at all",
                                "2" = "Slightly well",
                                "3" = "Moderately well",
```

```
                                 "4" = "Very well",
                                 "5" = "Extremely well"),
contributionunderstanding_Video = fct_recode
(contributionunderstanding_Video,
                                 "1" = "Not well at all",
                                 "2" = "Slightly well",
                                 "3" = "Moderately well",
                                 "4" = "Very well",
                                 "5" = "Extremely well"),
knowledge_Stories = fct_recode(knowledge_Stories,
                        "1" = "Strongly disagree",
                        "2" = "Somewhat disagree",
                        "3" = "Neither agree nor disagree",
                        "4" = "Somewhat agree",
                        "5" = "Strongly agree"),
knowledge_Video = fct_recode(knowledge_Video,
                        "1" = "Strongly disagree",
                        "2" = "Somewhat disagree",
                        "3" = "Neither agree nor disagree",
                        "4" = "Somewhat agree",
                        "5" = "Strongly agree"),
pace_Stories = fct_recode(pace_Stories,
                        "1" = "Strongly disagree",
                        "2" = "Somewhat disagree",
                        "3" = "Neither agree nor disagree",
                        "4" = "Somewhat agree",
                        "5" = "Strongly agree"),
pace_Video = fct_recode(pace_Video,
                   "1" = "Strongly disagree",
                   "2" = "Somewhat disagree",
                   "3" = "Neither agree nor disagree",
                   "4" = "Somewhat agree",
                   "5" = "Strongly agree"),
compelling_Stories = fct_recode(compelling_Stories,
                         "1" = "Strongly disagree",
                         "2" = "Somewhat disagree",
                         "3" = "Neither agree nor disagree",
                         "4" = "Somewhat agree",
                         "5" = "Strongly agree"),
compelling_Video = fct_recode(compelling_Video,
                         "1" = "Strongly disagree",
                         "2" = "Somewhat disagree",
                         "3" = "Neither agree nor disagree",
```

```
                         "4" = "Somewhat agree",
                         "5" = "Strongly agree"),
boring_Stories = fct_recode(boring_Stories,
                         "1" = "Strongly disagree",
                         "2" = "Somewhat disagree",
                         "3" = "Neither agree nor disagree",
                         "4" = "Somewhat agree",
                         "5" = "Strongly agree"),
boring_Video = fct_recode(boring_Video,
                       "1" = "Strongly disagree",
                       "2" = "Somewhat disagree",
                       "3" = "Neither agree nor disagree",
                       "4" = "Somewhat agree",
                       "5" = "Strongly agree")
         )

# Add unique ID for each user
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(User = rownames(survey_quant_tidy))

# Convert to a categorical factor
survey_quant_tidy$User = factor(survey_quant_tidy$User)
```

```
# Read in a data file with all qualitative survey data
survey_qual_raw = read_csv("survey_qual.csv")

# Tidy qualitative survey data
survey_qual_tidy <- survey_qual_raw %>%
  dplyr::select(Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q16) %>%
  dplyr::rename("area_of_study" = "Q4",
         "stories_familiarity" = "Q5",
         "technical_expertise" = "Q6",
         "attention" = "Q7",
         "attention_comments" = "Q8",
         "learning" = "Q9",
         "learning_comments" = "Q10",
         "final_comments" = "Q16"
  ) %>%
  slice(3:n()) %>%
  filter(!is.na(area_of_study)) %>%
  mutate(stories_familiarity = fct_recode(stories_familiarity,
                                        "1" = "Not familiar at all",
                                        "2" = "Slightly familiar",
```

```r
                                                          "3" = "Moderately familiar",
                                                          "4" = "Very familiar",
                                                          "5" = "Extremely familiar")
          )

# Add unique ID for each user
survey_qual_tidy <- survey_qual_tidy %>%
  mutate(User = rownames(survey_qual_tidy))

# Convert to a categorical factor
survey_qual_tidy$User = factor(survey_qual_tidy$User)
```

```r
# From wide to long
library(tidyr)
library(readr)
survey_quant_long <- survey_quant_tidy %>%
  gather(key = Medium, contributionunderstanding_Stories:boring_Video,
         value = response) %>%
  separate(Medium, into = c("temp", "Medium"),sep = "\\_") %>%
  spread(temp, response) %>%
  # Convert variables of characters to integers so that we can
  # perform analyses on them
  type_convert()

# Convert to categorical factors
survey_quant_long$User = factor(survey_quant_long$User)
survey_quant_long$Medium = factor(survey_quant_long$Medium)
```

```r
# Ordinal variable analysis of Likert scale response (1-5)
# VARIABLE: Understanding of key contributions
# Descriptive statistics by Medium
library(plyr)
ddply(survey_quant_long, ~ Medium, function(data)
summary(data$contributionunderstanding))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories    3       4      4 4.136364    4.75    5
## 2   Video    3       3      4 3.772727    4.00    5
```

```r
ddply(survey_quant_long, ~ Medium, summarise,
      contributionunderstanding.mean=mean(contributionunderstanding),
      contributionunderstanding.sd=sd(contributionunderstanding))
```

```
##    Medium contributionunderstanding.mean contributionunderstanding.sd
## 1 Stories                       4.136364                     0.6396021
## 2   Video                       3.772727                     0.6119304
```

```
# Nonparametric Wilcoxon signed-rank test for ordinal variables
library(coin)
wilcoxsign_test(contributionunderstanding ~ Medium | User,
    data=survey_quant_long, distribution="exact")
```

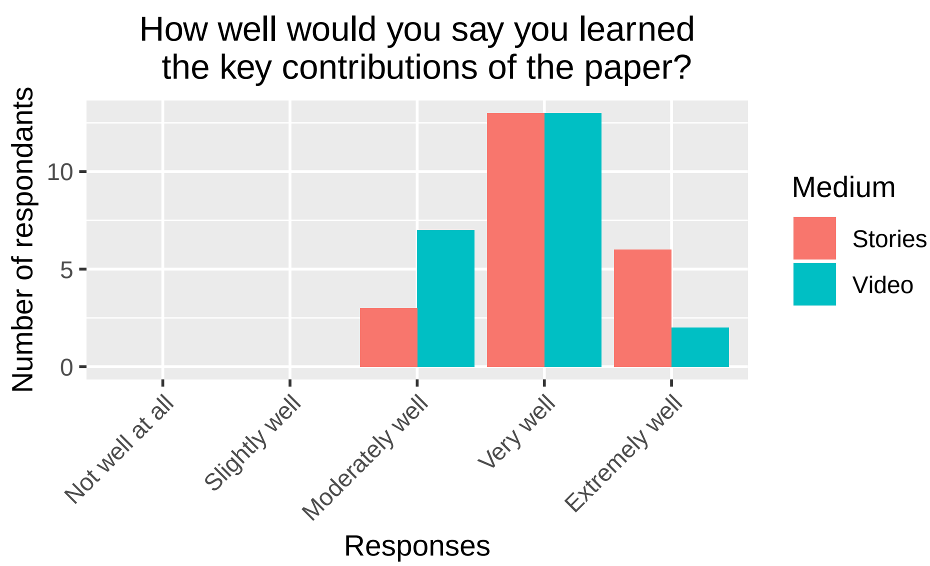
```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 2.1498, p-value = 0.03809
## alternative hypothesis: true mu is not equal to 0
```

```
# Likert scale grouped histograms
library(ggplot2)
survey_quant_long <- survey_quant_long %>%
  mutate(contributionunderstanding = factor(contributionunderstanding,
                                            levels = c("1", "2", "3",
                                                        "4", "5")))
survey_quant_long$contributionunderstanding
```

```
##  [1] 4 4 5 3 4 3 3 3 4 4 5 4 4 3 4 3 5 4 4 4 5 4 5 4 4 4 4 4 3 5 4 4
## [36] 4 5 4 3 3 4 5 4 3
## Levels: 1 2 3 4 5
```

```
survey_quant_long %>%
  ggplot() +
  geom_bar(aes(contributionunderstanding, fill = Medium),
    position = "dodge") +
  scale_x_discrete(drop = F, labels = c("1" = "Not well at all",
                                        "2" = "Slightly well",
                                        "3" = "Moderately well",
                                        "4" = "Very well",
                                        "5" = "Extremely well")) +
  scale_fill_discrete(labels = c(stories = "Stories",
                                 video = "Video")) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ggtitle("How well would you say you learned
```

```
  the key contributions of the paper?") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# VARIABLE: Improvement in knowledge of paper's subject matter
# Descriptive statistics by Medium
library(plyr)
ddply(survey_quant_long, ~ Medium, function(data)
summary(data$knowledge))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories    3       4      5 4.454545       5    5
## 2   Video    3       4      5 4.500000       5    5
```

```
ddply(survey_quant_long, ~ Medium, summarise,
    knowledge.mean=mean(knowledge), knowledge.sd=sd(knowledge))
```

```
##    Medium knowledge.mean knowledge.sd
## 1 Stories       4.454545    0.6709817
## 2   Video       4.500000    0.5976143
```

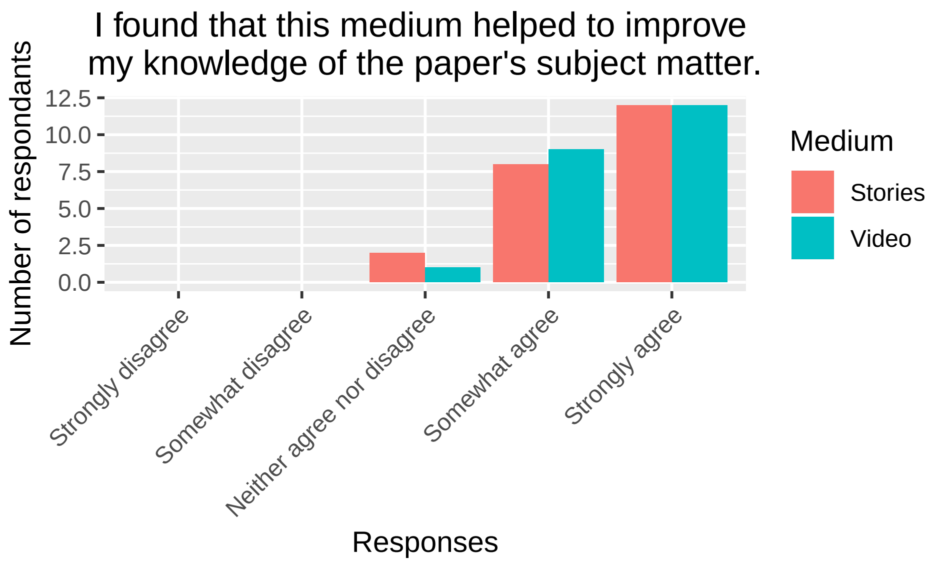
```
# Nonparametric Wilcoxon signed-rank test for ordinal variables
library(coin)
wilcoxsign_test(knowledge ~ Medium | User, data=survey_quant_long,
    distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = -0.062704, p-value = 1
## alternative hypothesis: true mu is not equal to 0
```

```
# Likert scale grouped histograms
library(ggplot2)
survey_quant_long <- survey_quant_long %>%
  mutate(knowledge = factor(knowledge, levels = c("1", "2", "3", "4",
  "5")))
survey_quant_long$knowledge
```

```
##  [1] 4 4 5 4 4 4 4 4 3 3 5 5 4 4 5 5 5 4 5 5 5 4 5 5 4 5 4 5 5 5 3 5
## [36] 5 5 5 5 5 4 4 5 5
## Levels: 1 2 3 4 5
```

```
survey_quant_long %>%
  ggplot() +
  geom_bar(aes(knowledge, fill = Medium), position = "dodge") +
  scale_x_discrete(drop = F,
  labels = c(
              "1" = "Strongly disagree",
              "2" = "Somewhat disagree",
              "3" = "Neither agree nor disagree",
              "4" = "Somewhat agree",
              "5" = "Strongly agree")) +
  scale_fill_discrete(labels = c(stories = "Stories",
                                  video = "Video")) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ggtitle("I found that this medium helped to improve
my knowledge of the paper's subject matter.") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
# VARIABLE: Pace of comprehension
# Descriptive statistics by Medium
library(plyr)
ddply(survey_quant_long, ~ Medium, function(data) summary(data$pace))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories    3    4.25      5 4.681818       5    5
## 2   Video    1    3.00      4 3.500000       4    5
```

```r
ddply(survey_quant_long, ~ Medium, summarise, pace.mean=mean(pace),
    pace.sd=sd(pace))
```

```
##    Medium pace.mean   pace.sd
## 1 Stories  4.681818 0.5679004
## 2   Video  3.500000 1.0578505
```

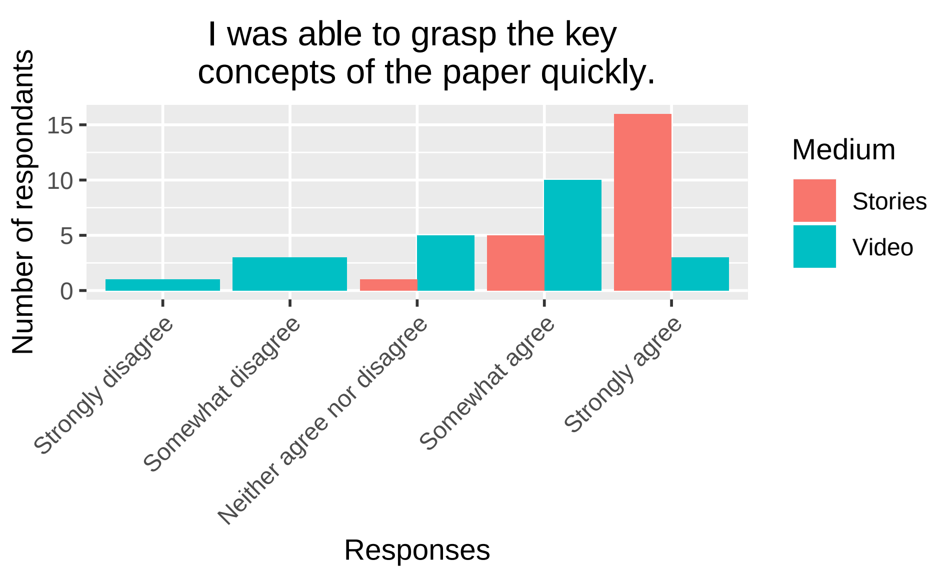
```r
# Nonparametric Wilcoxon signed-rank test for ordinal variables
library(coin)
wilcoxsign_test(pace ~ Medium | User, data=survey_quant_long,
    distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 3.4677, p-value = 0.0003166
## alternative hypothesis: true mu is not equal to 0
```

```r
# Likert scale grouped histograms
survey_quant_long <- survey_quant_long %>%
  mutate(pace = factor(pace, levels = c("1", "2", "3", "4", "5")))
survey_quant_long$pace
```

```
##  [1] 4 3 5 4 5 1 5 2 4 2 5 4 4 4 5 4 5 4 5 4 5 3 5 4 5 4 5 2 5 5 5 3
## [36] 4 4 3 4 5 5 5 5 3
## Levels: 1 2 3 4 5
```

```r
survey_quant_long %>%
  ggplot() +
  geom_bar(aes(pace, fill = Medium), position = "dodge") +
  labs(fill = "Medium") +
  scale_x_discrete(drop = F,
  labels = c(
              "1" = "Strongly disagree",
              "2" = "Somewhat disagree",
              "3" = "Neither agree nor disagree",
              "4" = "Somewhat agree",
              "5" = "Strongly agree")) +
  scale_fill_discrete(labels = c(stories = "Stories",
                                 video = "Video")) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ggtitle("I was able to grasp the key
  concepts of the paper quickly.") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
# VARIABLE: Compelling rating
# Descriptive statistics by Medium
library(plyr)
ddply(survey_quant_long, ~ Medium, function(data)
summary(data$compelling))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories    3       4      5 4.454545       5    5
## 2   Video    2       4      4 4.045455       5    5
```

```r
ddply(survey_quant_long, ~ Medium, summarise,
    compelling.mean=mean(compelling),
    compelling.sd=sd(compelling))
```

```
##    Medium compelling.mean compelling.sd
## 1 Stories        4.454545     0.6709817
## 2   Video        4.045455     1.0455016
```

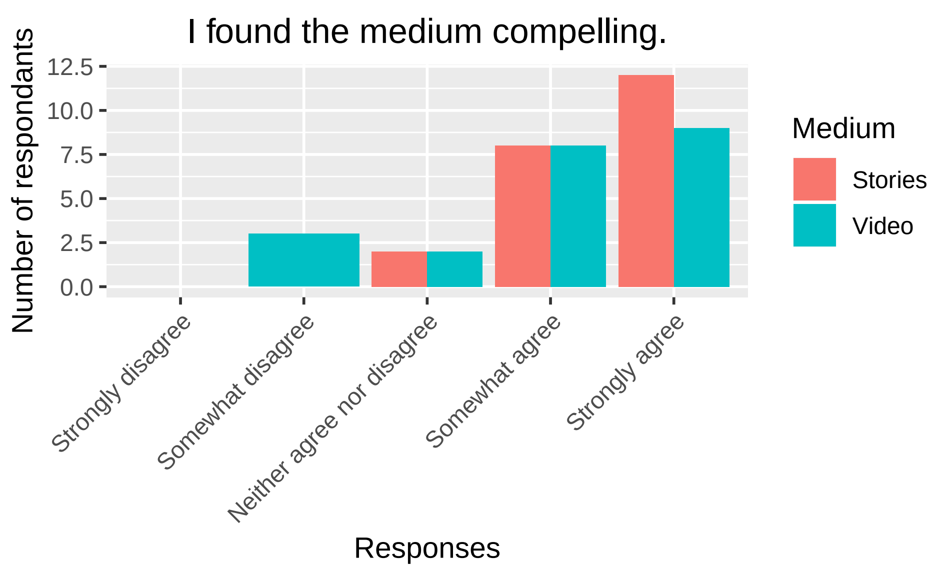
```r
# Nonparametric Wilcoxon signed-rank test for ordinal variables
library(coin)
wilcoxsign_test(compelling ~ Medium | User, data=survey_quant_long,
    distribution="exact")
```

```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = 1.0702, p-value = 0.2998
## alternative hypothesis: true mu is not equal to 0
```

```r
# Likert scale grouped histograms
library(ggplot2)
survey_quant_long <- survey_quant_long %>%
  mutate(compelling = factor(compelling, levels = c("1", "2", "3", "4",
                                                     "5")))
survey_quant_long$compelling
```

```
##  [1] 4 5 5 2 5 3 3 5 5 4 5 4 4 4 5 2 5 4 5 5 5 2 5 5 5 5 3 3 4 5 4 4
## [36] 5 5 4 4 5 4 5 4 4
## Levels: 1 2 3 4 5
```

```r
survey_quant_long %>%
  ggplot() +
  geom_bar(aes(compelling, fill = Medium), position = "dodge") +
  labs(fill = "Medium") +
  scale_x_discrete(drop = F,
  labels = c("1" = "Strongly disagree",
             "2" = "Somewhat disagree",
             "3" = "Neither agree nor disagree",
             "4" = "Somewhat agree",
             "5" = "Strongly agree")) +
  scale_fill_discrete(labels = c(stories = "Stories",
                                 video = "Video")) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ggtitle("I found the medium compelling.") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
# VARIABLE: Boring rating
# Descriptive statistics by Medium
library(plyr)
ddply(survey_quant_long, ~ Medium, function(data)
summary(data$boring))
```

```
##    Medium Min. 1st Qu. Median     Mean 3rd Qu. Max.
## 1 Stories    1       2      2 2.181818    2.75    4
## 2   Video    1       2      3 2.863636    4.00    5
```

```r
ddply(survey_quant_long, ~ Medium, summarise, boring.mean=mean(boring),
      boring.sd=sd(boring))
```

```
##    Medium boring.mean boring.sd
## 1 Stories    2.181818 0.8528029
## 2   Video    2.863636 1.2069424
```

```r
# Nonparametric Wilcoxon signed-rank test for ordinal variables
library(coin)
wilcoxsign_test(boring ~ Medium | User, data=survey_quant_long,
    distribution="exact")
```

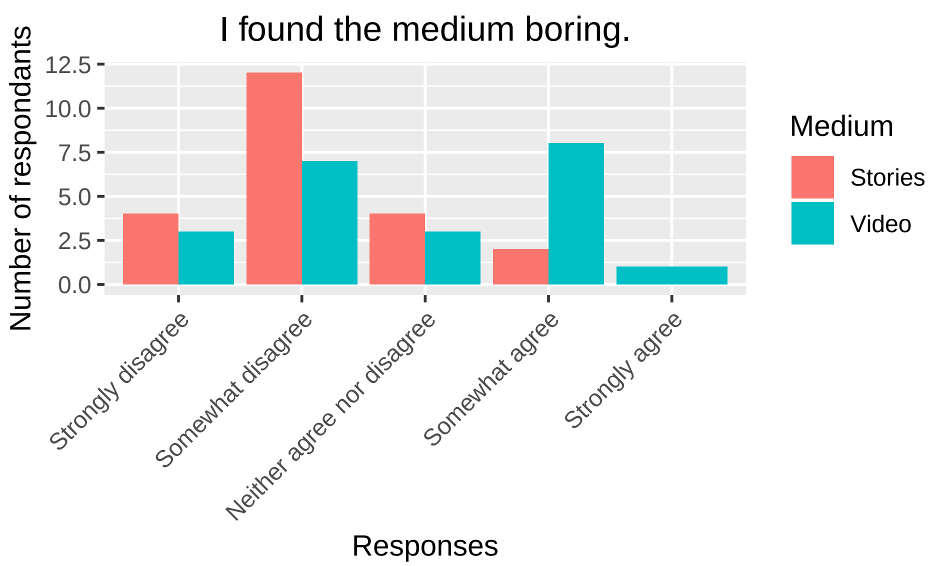
```
##
##  Exact Wilcoxon-Pratt Signed-Rank Test
##
## data:  y by x (pos, neg)
##   stratified by block
## Z = -2.1166, p-value = 0.03711
## alternative hypothesis: true mu is not equal to 0
```

```r
# Likert scale grouped histograms
survey_quant_long <- survey_quant_long %>%
  mutate(boring = factor(boring, levels = c("1", "2", "3", "4", "5")))
survey_quant_long$boring
```

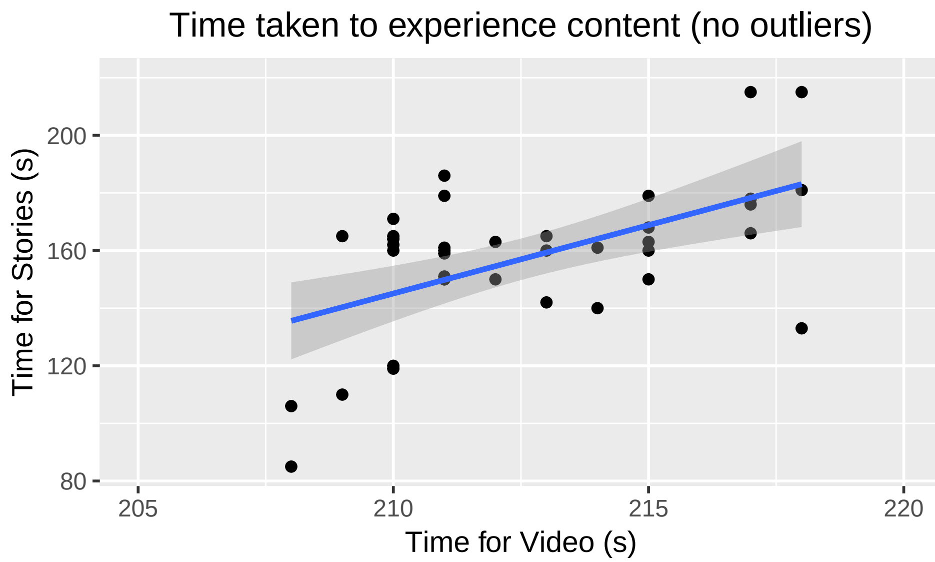
```
##  [1] 2 1 2 4 2 4 3 2 2 2 1 1 3 3 2 4 1 4 4 4 1 3 1 1 2 2 3 3 3 2 2 4
## [36] 2 2 4 4 5 2 2 2 4
## Levels: 1 2 3 4 5
```

```r
survey_quant_long %>%
  ggplot() +
  geom_bar(aes(boring, fill = Medium), position = "dodge") +
  labs(fill = "Medium") +
  scale_x_discrete(drop = F,
  labels = c("1" = "Strongly disagree",
            "2" = "Somewhat disagree",
            "3" = "Neither agree nor disagree",
            "4" = "Somewhat agree",
            "5" = "Strongly agree")) +
  scale_fill_discrete(labels = c(stories = "Stories",
                                 video = "Video")) +
  xlab("Responses") +
```

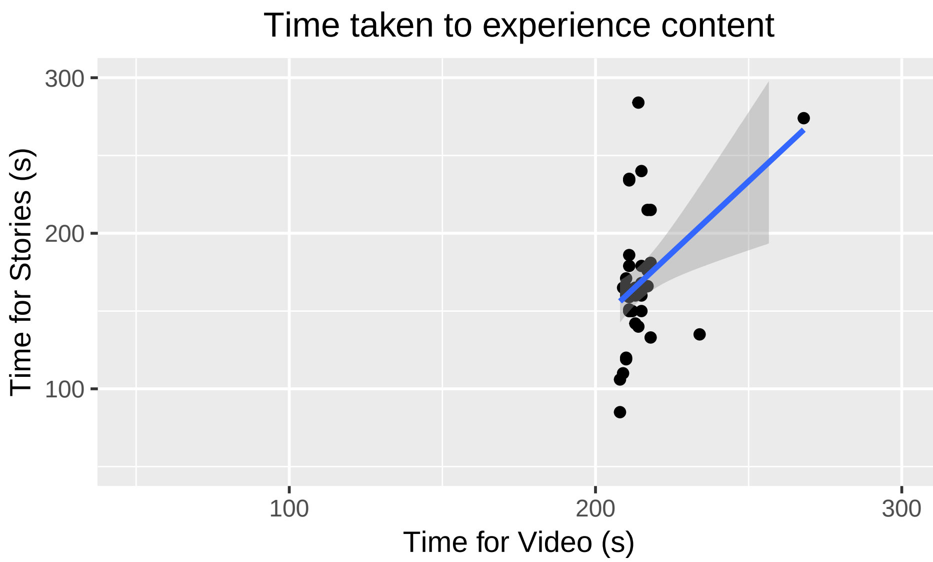
```r
  ylab("Number of respondants") +
  ggtitle("I found the medium boring.") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```



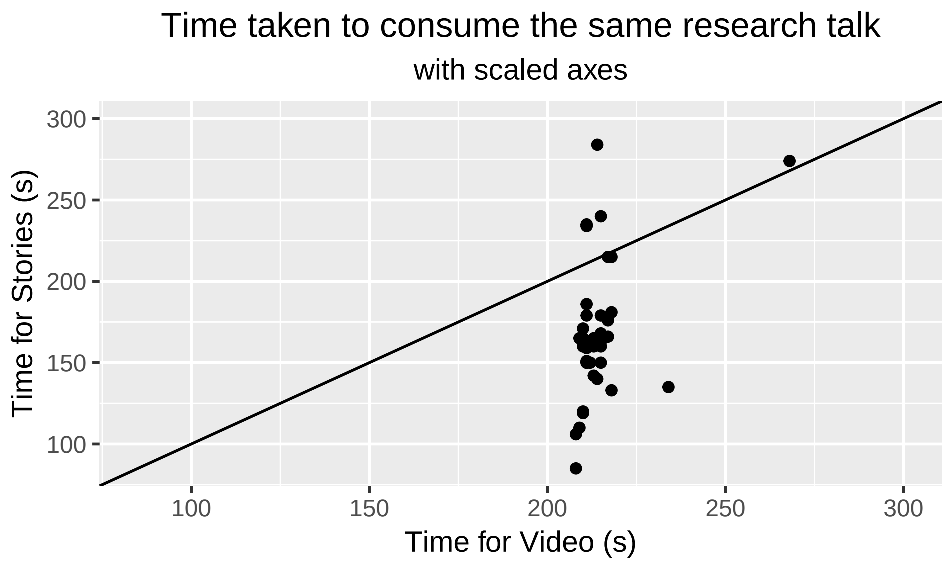

```r
# Linear model comparing time for both mediums
# No outliers
ggplot() +
  geom_point(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time)) +
  stat_smooth(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time), method = "lm") +
  xlim(205,220) +
  ylim(85,220) +
  xlab("Time for Video (s)") +
  ylab("Time for Stories (s)") +
  ggtitle("Time taken to experience content (no outliers)") +
  theme(plot.title = element_text(hjust = 0.5))
```



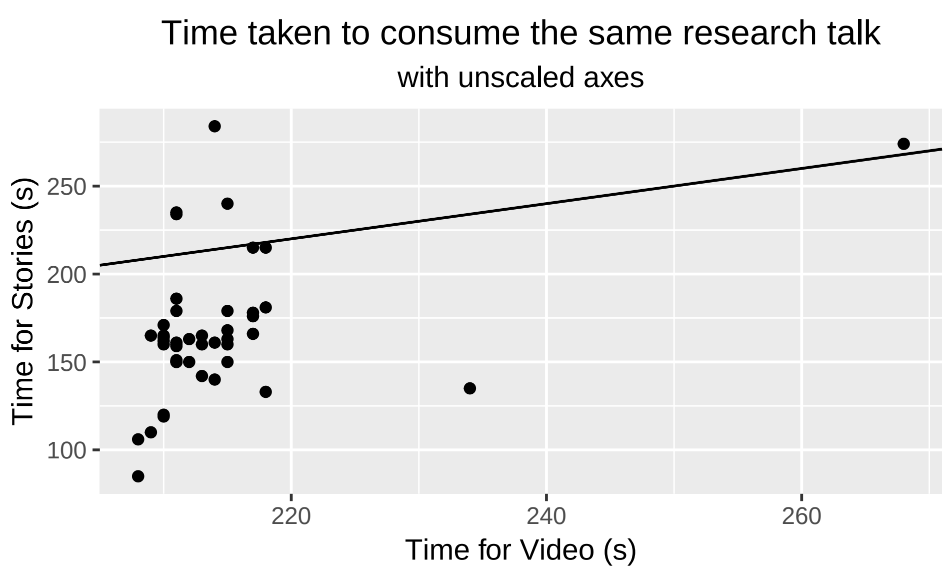

```
# All data
ggplot() +
  geom_point(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time)) +
  stat_smooth(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time), method = "lm") +
  xlab("Time for Video (s)") +
  ylab("Time for Stories (s)") +
  ggtitle("Time taken to experience content") +
  theme(plot.title = element_text(hjust = 0.5)) +
xlim(50,300) +
  ylim(50,300)
```




```r
# Scatterplots with line y = x to illustrate which users
# consumed content faster in a certain medium
#
# Scaled
ggplot() +
  geom_point(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time)) +
  geom_abline() +
  labs(x = "Time for Video (s)",
       y = "Time for Stories (s)",
       title = "Time taken to consume the same research talk",
       subtitle = "with scaled axes") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  xlim(85,300) +
  ylim(85,300)
```



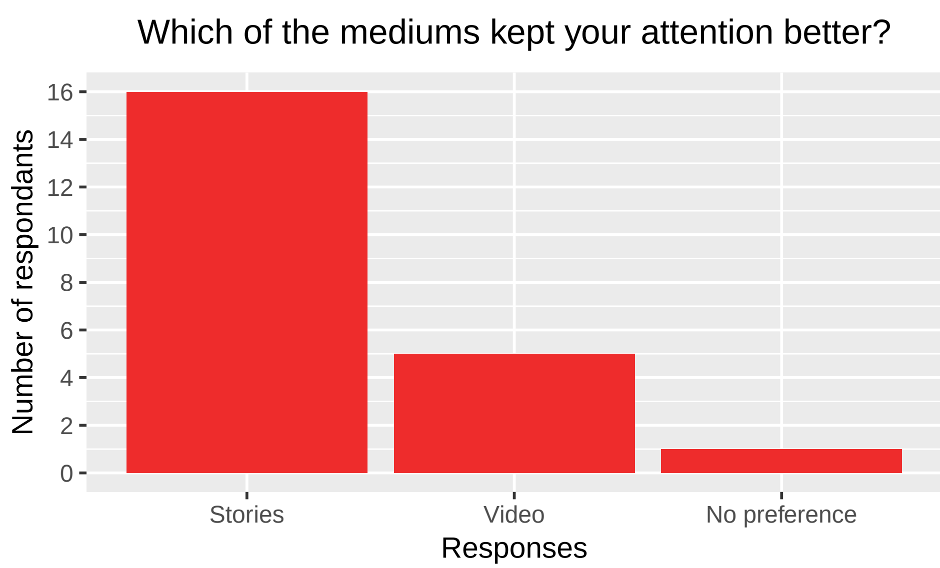

```r
# Unscaled
ggplot() +
  geom_point(aes(studyrecordings[studyrecordings$Medium ==
 "Video",]$Time, studyrecordings[studyrecordings$Medium ==
 "Stories",]$Time)) +
  geom_abline() +
  labs(x = "Time for Video (s)",
       y = "Time for Stories (s)",
       title = "Time taken to consume the same research talk",
       subtitle = "with unscaled axes") +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



```r
# Categorical variable analysis
# VARIABLE: Attention
# Histograms
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(attention = factor(attention, levels = c("Stories", "Video",
  "No preference")))
survey_quant_tidy$attention
```

```
##  [1] Stories       Stories       Stories       Video         Stories
##  [6] Stories       Stories       Stories       Stories       Stories
## [11] Stories       Video         Stories       Video         Video
## [16] Stories       Stories       Video         Stories       No pref
## [21] Stories       Stories
## Levels: Stories Video No preference
```

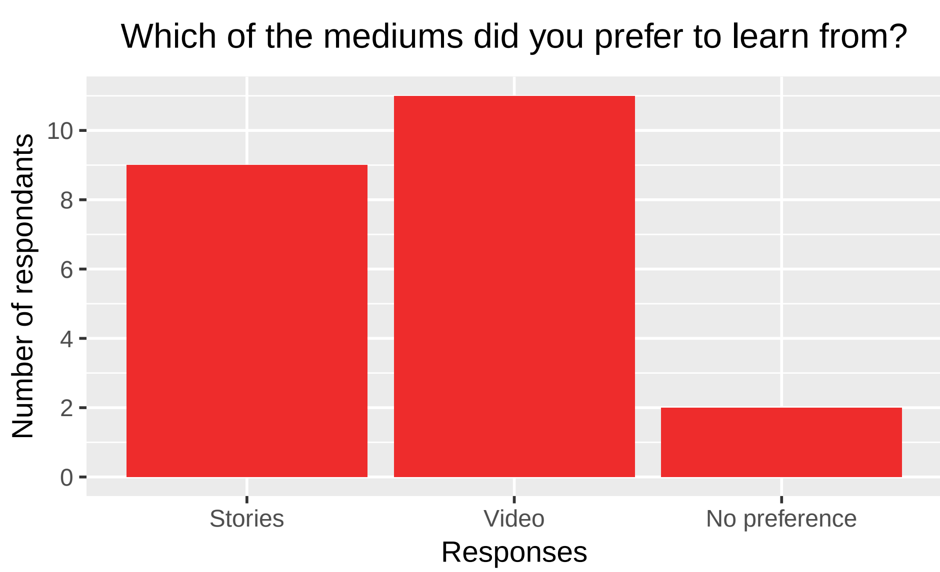
```r
survey_quant_tidy %>%
  ggplot() +
  geom_bar(aes(attention), position = "dodge", fill = "firebrick2") +
  scale_x_discrete(drop = F) +
  xlab("Responses") +
  ylab("Number of respondants") +
  scale_y_continuous(breaks = c(0,2,4,6,8,10,12,14,16)) +
  ggtitle(expression(paste(
        "Which of the mediums kept your attention better?"))) +
  theme(plot.title = element_text(hjust = 0.5))
```



```r
# VARIABLE: Pleasant to learn from
# Histograms
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(learning = factor(learning, levels = c("Stories", "Video"
  , "No preference")))
survey_quant_tidy$learning
```

```
##  [1] Stories       Video         Video         Video         Video
##  [6] Video         Video         No preference Stories       Stories
## [11] Stories       Video         Video         Stories       Video
## [16] Video         Stories       Video         Stories       Stories
## [21] No preference Stories
## Levels: Stories Video No preference
```

```r
survey_quant_tidy %>%
  ggplot() +
  geom_bar(aes(learning), position = "dodge", fill = "firebrick2") +
  scale_x_discrete(drop = F) +
  xlab("Responses") +
  ylab("Number of respondants") +
  scale_y_continuous(breaks = c(0,2,4,6,8,10,12)) +
  ggtitle(expression(paste(
            "Which of the mediums did you prefer to learn from?"))) +
  theme(plot.title = element_text(hjust = 0.5))
```

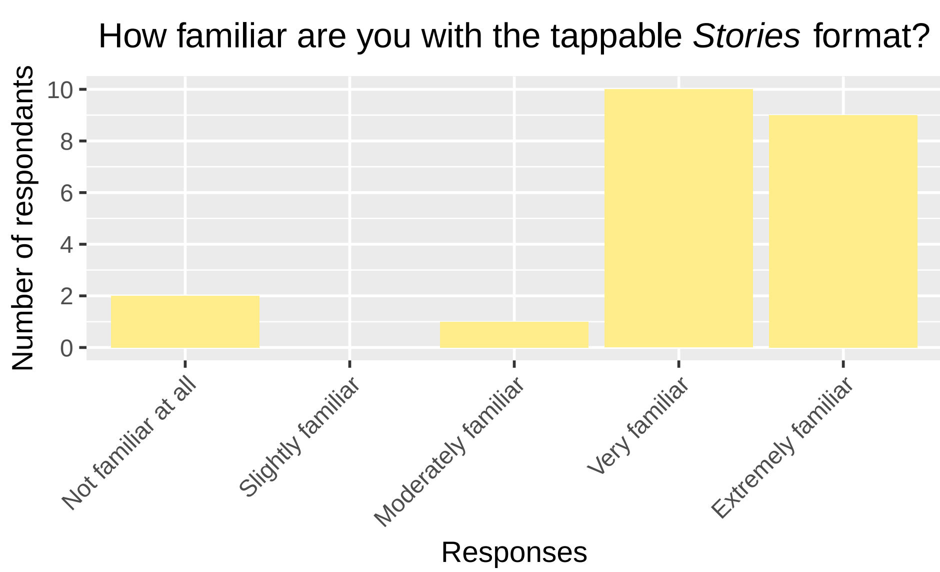



```r
# Demographics data (categorical)
# VARIABLE: Familiarity with the Stories medium
# Histograms
```

```r
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(stories_familiarity = factor(stories_familiarity,
                                      levels = c("1", "2", "3", "4",
                                                 "5")))
survey_quant_tidy$stories_familiarity
```

```
##  [1] 5 4 4 4 5 4 4 5 3 4 5 4 4 5 4 5 5 5 5 1 1 4
## Levels: 1 2 3 4 5
```

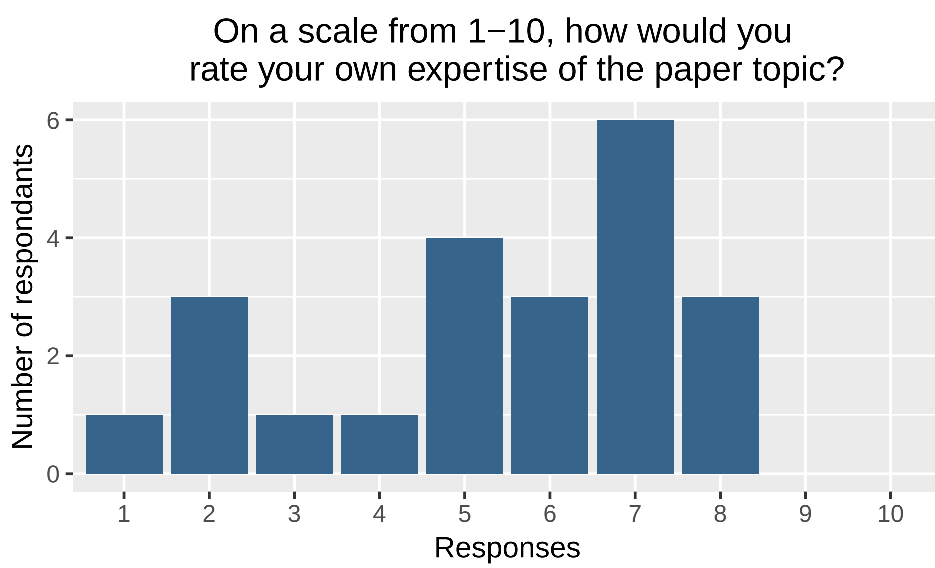
```r
survey_quant_tidy %>%
  ggplot() +
  geom_bar(aes(stories_familiarity), position = "dodge",
           fill = "lightgoldenrod1") +
  scale_x_discrete(drop = F, labels = c("1" = "Not familiar at all",
                                        "2" = "Slightly familiar",
                                        "3" = "Moderately familiar",
                                        "4" = "Very familiar",
                                        "5" = "Extremely familiar")) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ylim(0,10) +
  scale_y_continuous(breaks = c(0,2,4,6,8,10)) +
  ggtitle(expression(paste("How familiar are you with the tappable ",
  italic("Stories"),
  " format?"))) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
# VARIABLE: Technical expertise
# Histograms
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(technical_expertise = factor(technical_expertise,
                                      levels = c("1", "2", "3", "4",
                                                 "5", "6", "7", "8",
                                                 "9", "10")))
survey_quant_tidy$technical_expertise
```

```
##  [1] 6 7 8 5 8 5 6 5 2 6 7 7 7 7 3 8 2 4 7 2 1 5
## Levels: 1 2 3 4 5 6 7 8 9 10
```

```r
survey_quant_tidy %>%
  ggplot() +
  geom_bar(aes(technical_expertise), position = "dodge",
              fill = "steelblue4") +
  scale_x_discrete(drop = F) +
  xlab("Responses") +
  ylab("Number of respondants") +
  ggtitle("On a scale from 1-10, how would you
  rate your own expertise of the paper topic?") +
  theme(plot.title = element_text(hjust = 0.5))
```



B Data

Data

Ieva Burk

```
# DATA SET 1: Performance measures
```

```
performance <- read_csv("studyrecordings.csv")
knitr::kable(performance)
```

User	Academic_paper	Order	Medium	Time	Errors_made	Expressions_of_satisfaction
1	AI	1	Stories	160	0	2
2	AI	1	Stories	171	0	1
3	AI	1	Stories	151	1	0
4	AI	1	Stories	165	0	0
5	AI	1	Stories	161	1	0
6	AI	1	Stories	85	1	0
7	AI	1	Stories	120	1	1
8	AI	1	Stories	234	0	1
9	AI	1	Stories	160	0	2
10	AI	1	Stories	162	0	1
11	AI	1	Stories	159	0	1
12	AI	1	Stories	150	0	0
13	AI	2	Stories	135	1	1
14	AI	2	Stories	160	0	0
15	AI	1	Stories	161	0	0
16	AI	2	Stories	186	0	0
17	AI	2	Stories	150	0	4
18	AI	2	Stories	179	0	1
19	AI	1	Stories	160	0	0
20	AI	1	Stories	106	1	0
21	AI	2	Stories	119	0	0
22	AI	1	Stories	110	0	0
1	AI	1	Video	215	0	1
2	AI	1	Video	210	0	0
3	AI	1	Video	211	0	0
4	AI	1	Video	210	0	1
5	AI	1	Video	211	0	0
6	AI	1	Video	208	0	0
7	AI	1	Video	210	0	0
8	AI	1	Video	211	0	1
9	AI	1	Video	211	0	0
10	AI	1	Video	210	0	0
11	AI	1	Video	211	0	1
12	AI	1	Video	211	0	1
13	AI	2	Video	234	0	0
14	AI	2	Video	213	0	1
15	AI	1	Video	214	0	0
16	AI	2	Video	211	0	0
17	AI	2	Video	211	0	1
18	AI	2	Video	211	0	0
19	AI	1	Video	210	0	0
20	AI	1	Video	208	0	1
21	AI	2	Video	210	0	0

User	Academic_paper	Order	Medium	Time	Errors_made	Expressions_of_satisfaction
22	AI	1	Video	209	0	0
1	CV	1	Stories	168	0	1
2	CV	1	Stories	176	0	0
3	CV	1	Stories	140	0	0
4	CV	1	Stories	165	0	0
5	CV	1	Stories	178	0	0
6	CV	1	Stories	133	0	1
7	CV	1	Stories	163	2	0
8	CV	1	Stories	274	2	1
9	CV	1	Stories	164	0	0
10	CV	1	Stories	166	0	0
11	CV	1	Stories	165	0	1
12	CV	1	Stories	284	0	3
13	CV	2	Stories	240	2	1
14	CV	2	Stories	150	0	0
15	CV	1	Stories	163	0	0
16	CV	2	Stories	181	1	0
17	CV	2	Stories	215	0	1
18	CV	2	Stories	235	0	2
19	CV	1	Stories	142	0	0
20	CV	1	Stories	215	0	0
21	CV	2	Stories	179	0	0
22	CV	1	Stories	150	0	0
1	CV	1	Video	215	0	0
2	CV	1	Video	217	0	0
3	CV	1	Video	214	0	0
4	CV	1	Video	209	0	0
5	CV	1	Video	217	0	0
6	CV	1	Video	218	0	0
7	CV	1	Video	215	0	1
8	CV	1	Video	268	1	1
9	CV	1	Video	210	0	0
10	CV	1	Video	217	0	0
11	CV	1	Video	213	0	1
12	CV	1	Video	214	0	0
13	CV	2	Video	215	0	0
14	CV	2	Video	215	0	1
15	CV	1	Video	212	0	0
16	CV	2	Video	218	0	0
17	CV	2	Video	218	0	2
18	CV	2	Video	211	0	0
19	CV	1	Video	213	0	0
20	CV	1	Video	217	0	0
21	CV	2	Video	215	0	0
22	CV	1	Video	212	0	0

```
# DATA SET 2: Subjective measures
```

```
# Read in a data file with all quantitative survey data
survey_quant_raw = read_csv("survey_quant.csv")
```

```r
# Tidy quantitative survey data
library(tidyverse)
require(MASS)
require(dplyr)

survey_quant_tidy <- survey_quant_raw %>%
  dplyr::select(Q3, Q4, Q5, Q6, Q7, Q9, Q11_1, Q11_2, Q12_1, Q12_2, Q13_1, Q13_2, Q14_1,
              Q14_2, Q15_1, Q15_2) %>%
  dplyr::rename("affiliation" = "Q3",
         "area_of_study" = "Q4",
         "stories_familiarity" = "Q5",
         "technical_expertise" = "Q6",
         "attention" = "Q7",
         "learning" = "Q9",
         "contributionunderstanding_Stories" = "Q11_1",
         "contributionunderstanding_Video" = "Q11_2",
         "knowledge_Stories" = "Q12_1",
         "knowledge_Video" = "Q12_2",
         "pace_Stories" = "Q13_1",
         "pace_Video" = "Q13_2",
         "compelling_Stories" = "Q14_1",
         "compelling_Video" = "Q14_2",
         "boring_Stories" = "Q15_1",
         "boring_Video" = "Q15_2"
         ) %>%
  mutate(affiliation = recode(affiliation, Other = "Rice Graduate '18")) %>%
  slice(3:n()) %>%
  filter(!is.na(affiliation)) %>%
  mutate(stories_familiarity = factor(stories_familiarity,
                                      levels = c("Not familiar at all",
                                                 "Slightly familiar",
                                                 "Moderately familiar",
                                                 "Very familiar",
                                                 "Extremely familiar")),
         contributionunderstanding_Stories = factor(contributionunderstanding_Stories,
                                                    levels = c("Not well at all",
                                                               "Slightly well",
                                                               "Moderately well",
                                                               "Very well",
                                                               "Extremely well")),
         contributionunderstanding_Video = factor(contributionunderstanding_Video,
                                                  levels = c("Not well at all",
                                                             "Slightly well",
                                                             "Moderately well",
                                                             "Very well",
                                                             "Extremely well")),
         knowledge_Stories = factor(knowledge_Stories,
                                    levels = c("Strongly disagree",
                                               "Somewhat disagree",
                                               "Neither agree nor disagree",
                                               "Somewhat agree",
                                               "Strongly agree")),
         knowledge_Video = factor(knowledge_Video,
```

```
                                  levels = c("Strongly disagree",
                                             "Somewhat disagree",
                                             "Neither agree nor disagree",
                                             "Somewhat agree",
                                             "Strongly agree")),
         pace_Stories = factor(pace_Stories,
                               levels = c("Strongly disagree",
                                          "Somewhat disagree",
                                          "Neither agree nor disagree",
                                          "Somewhat agree",
                                          "Strongly agree")),
         pace_Video = factor(pace_Video,
                             levels = c("Strongly disagree",
                                        "Somewhat disagree",
                                        "Neither agree nor disagree",
                                        "Somewhat agree",
                                        "Strongly agree")),
         compelling_Stories = factor(compelling_Stories,
                                     levels = c("Strongly disagree",
                                                "Somewhat disagree",
                                                "Neither agree nor disagree",
                                                "Somewhat agree",
                                                "Strongly agree")),
         compelling_Video = factor(compelling_Video,
                                   levels = c("Strongly disagree",
                                                "Somewhat disagree",
                                                "Neither agree nor disagree",
                                                "Somewhat agree",
                                                "Strongly agree")),
         boring_Stories = factor(boring_Stories,
                                 levels = c("Strongly disagree",
                                                "Somewhat disagree",
                                                "Neither agree nor disagree",
                                                "Somewhat agree",
                                                "Strongly agree")),
         boring_Video = factor(boring_Video,
                               levels = c("Strongly disagree",
                                                "Somewhat disagree",
                                                "Neither agree nor disagree",
                                                "Somewhat agree",
                                                "Strongly agree"))) %>%
  mutate(stories_familiarity = fct_recode(stories_familiarity,
                                          "1" = "Not familiar at all",
                                          "2" = "Slightly familiar",
                                          "3" = "Moderately familiar",
                                          "4" = "Very familiar",
                                          "5" = "Extremely familiar"),
         contributionunderstanding_Stories = fct_recode(contributionunderstanding_Stories,
                                                        "1" = "Not well at all",
                                                        "2" = "Slightly well",
                                                        "3" = "Moderately well",
                                                        "4" = "Very well",
                                                        "5" = "Extremely well"),
```

```
         contributionunderstanding_Video = fct_recode(contributionunderstanding_Video,
                                                  "1" = "Not well at all",
                                                  "2" = "Slightly well",
                                                  "3" = "Moderately well",
                                                  "4" = "Very well",
                                                  "5" = "Extremely well"),
         knowledge_Stories = fct_recode(knowledge_Stories,
                                 "1" = "Strongly disagree",
                                 "2" = "Somewhat disagree",
                                 "3" = "Neither agree nor disagree",
                                 "4" = "Somewhat agree",
                                 "5" = "Strongly agree"),
         knowledge_Video = fct_recode(knowledge_Video,
                                 "1" = "Strongly disagree",
                                 "2" = "Somewhat disagree",
                                 "3" = "Neither agree nor disagree",
                                 "4" = "Somewhat agree",
                                 "5" = "Strongly agree"),
         pace_Stories = fct_recode(pace_Stories,
                                 "1" = "Strongly disagree",
                                 "2" = "Somewhat disagree",
                                 "3" = "Neither agree nor disagree",
                                 "4" = "Somewhat agree",
                                 "5" = "Strongly agree"),
         pace_Video = fct_recode(pace_Video,
                            "1" = "Strongly disagree",
                            "2" = "Somewhat disagree",
                            "3" = "Neither agree nor disagree",
                            "4" = "Somewhat agree",
                            "5" = "Strongly agree"),
         compelling_Stories = fct_recode(compelling_Stories,
                                  "1" = "Strongly disagree",
                                  "2" = "Somewhat disagree",
                                  "3" = "Neither agree nor disagree",
                                  "4" = "Somewhat agree",
                                  "5" = "Strongly agree"),
         compelling_Video = fct_recode(compelling_Video,
                                  "1" = "Strongly disagree",
                                  "2" = "Somewhat disagree",
                                  "3" = "Neither agree nor disagree",
                                  "4" = "Somewhat agree",
                                  "5" = "Strongly agree"),
         boring_Stories = fct_recode(boring_Stories,
                                  "1" = "Strongly disagree",
                                  "2" = "Somewhat disagree",
                                  "3" = "Neither agree nor disagree",
                                  "4" = "Somewhat agree",
                                  "5" = "Strongly agree"),
         boring_Video = fct_recode(boring_Video,
                                "1" = "Strongly disagree",
                                "2" = "Somewhat disagree",
                                "3" = "Neither agree nor disagree",
                                "4" = "Somewhat agree",
```

```r
                                         "5" = "Strongly agree")
         )

# Add unique ID for each user
survey_quant_tidy <- survey_quant_tidy %>%
  mutate(User = rownames(survey_quant_tidy)) %>%
  dplyr::select(User, everything())

# Convert to a categorical factor
survey_quant_tidy$User = factor(survey_quant_tidy$User)

library(knitr)
library(kableExtra)
kable(survey_quant_tidy) %>%
  kable_styling(latex_options = c("scale_down")) %>%
  row_spec(0, angle = 90)
```

User	affiliation	area_of_study	stories_familiarity	technical_expertise	attention	learning	contributionunderstanding_Stories	contributionunderstanding_Video	knowledge_Stories	knowledge_Video	pace_Stories	pace_Video	compelling_Stories	compelling_Video	boring_Stories	boring_Video
1	Rice Graduate '18	Economics	5	6	Stories	Stories	4	3	5	5	5	3	4	4	2	4
2	5C Student	Economics + Engineering (3-2)	4	7	Stories	Video	4	4	4	5	5	4	5	5	2	2
3	5C Student	Economics + Engineering (3-2)	4	8	Stories	Video	4	4	4	5	5	2	3	3	3	3
4	5C Student	Mathematics,Computer Science	4	5	Video	Video	4	4	4	4	3	4	4	4	2	2
5	5C Student	Computer Science	5	8	Stories	Video	4	3	5	5	5	4	5	2	2	4
6	5C Student	Mathematics,Computer Science	4	5	Stories	Video	4	4	3	5	5	3	4	4	2	4
7	5C Student	Computer Science	4	6	Stories	Video	3	3	4	4	5	2	3	5	3	2
8	5C Student	Computer Science,Economics,Physics	5	5	Stories	No preference	5	4	5	4	5	3	5	2	1	3
9	5C Student	PPE,French	3	2	Stories	Stories	5	4	5	5	4	3	5	4	2	4
10	5C Student	Computer Science	4	6	Stories	Stories	4	3	4	4	5	1	5	3	2	4
11	5C Student	Computer Science	5	7	Stories	Stories	5	4	5	5	5	4	5	4	1	1
12	5C Faculty Member	Computer Science,Psychology,Neuroscience	4	7	Video	Video	4	4	4	4	4	3	4	5	2	1
13	5C Student	Computer Science	4	7	Stories	Video	4	4	3	3	4	2	5	4	2	2
14	5C Student	Science Management	5	7	Video	Stories	4	5	4	4	5	5	4	5	2	2
15	5C Student	Economics,Film Studies	4	3	Video	Video	3	5	5	5	5	5	4	5	3	2
16	5C Student	Computer Science,Economics	5	8	Stories	Video	4	4	5	5	5	4	5	5	4	4
17	5C Student	Computer Science,PPE	5	2	Stories	Stories	5	4	5	5	5	4	5	5	1	1
18	5C Student	Psychology	5	4	Video	Video	3	3	5	5	4	5	4	5	4	5
19	5C Student	Computer Science	5	7	Stories	Stories	4	3	4	4	4	4	4	4	3	3
20	5C Student	PPE	1	2	No preference	Stories	4	4	5	5	5	4	5	5	2	2
21	5C Student	Computer Science,Economics	1	1	Stories	No preference	5	4	5	4	5	4	5	4	1	4
22	5C Student	Chemistry	4	5	Stories	Stories	5	3	5	4	5	4	5	2	2	4

```r
# Long format conversion
library(tidyr)
survey_quant_long <- survey_quant_tidy %>%
  gather(key = Medium, contributionunderstanding_Stories:boring_Video,
         value = response) %>%
  separate(Medium, into = c("temp", "Medium"),sep = "\\_") %>%
  spread(temp, response) %>%
  # Convert variables of characters to integers so that we can perform analyses on them
  type_convert()

# Convert to a categorical factor
survey_quant_long$User = factor(survey_quant_long$User)
```

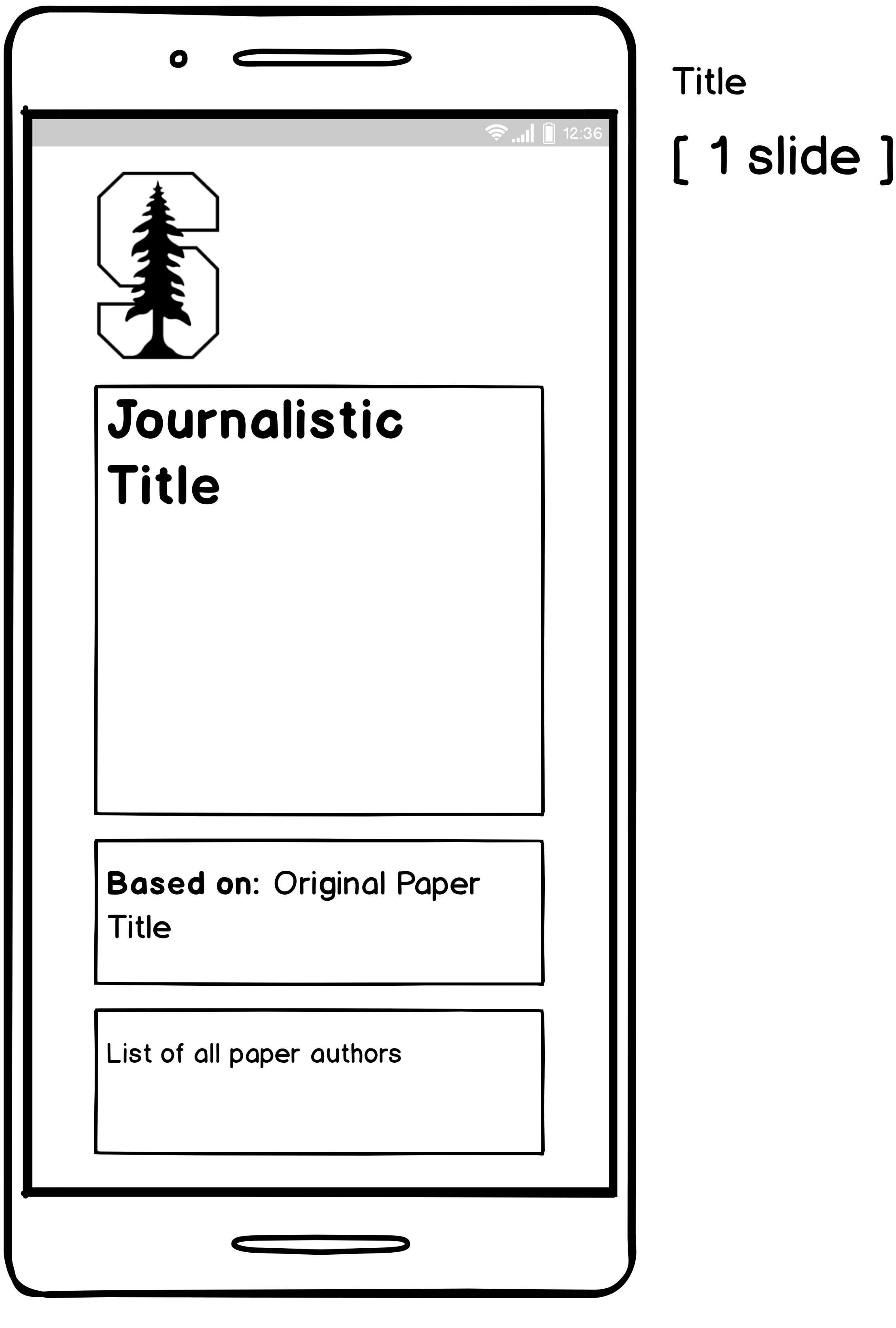
```r
survey_quant_long$Medium = factor(survey_quant_long$Medium)

library(knitr)
library(kableExtra)
kable(survey_quant_long) %>%
  kable_styling(latex_options = c("scale_down")) %>%
  row_spec(0, angle = 90)
```

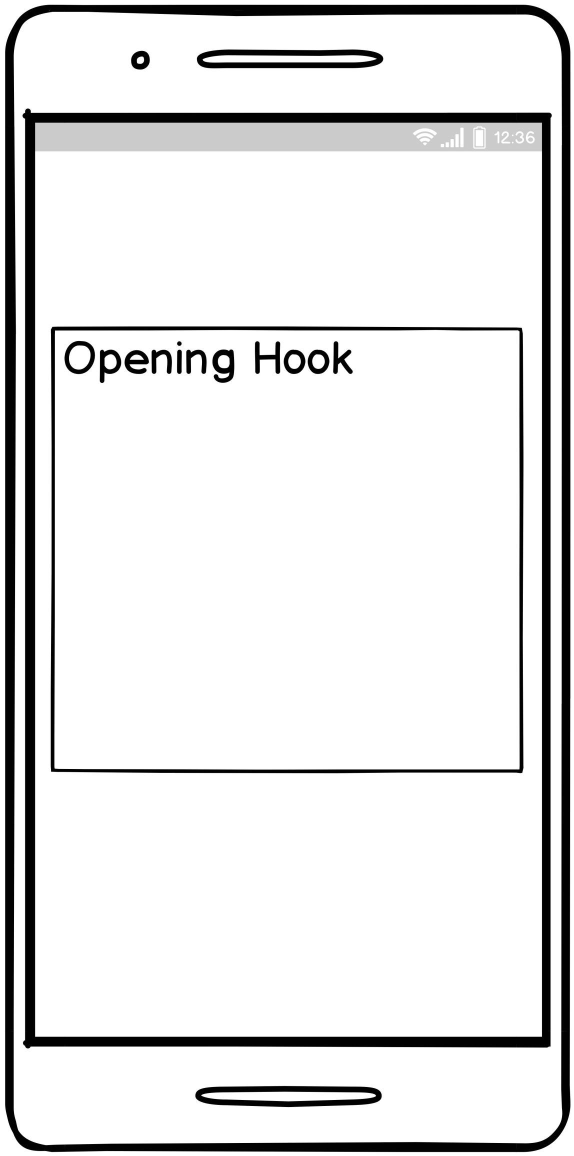
User	affiliation	area_of_study	stories_familiarity	technical_expertise	attention	learning	Medium	boring	compelling	contributionunderstanding	knowledge	pace
1	Rice Graduate '18	Economics	5	6	Stories	Stories	Stories	2	4	4	5	5
1	Rice Graduate '18	Economics	5	6	Stories	Stories	Video	4	4	3	5	3
10	5C Student	Computer Science	4	6	Stories	Stories	Stories	2	5	4	4	5
10	5C Student	Computer Science	4	6	Stories	Stories	Video	4	3	3	4	1
11	5C Student	Computer Science	5	7	Stories	Stories	Stories	1	5	5	5	5
11	5C Student	Computer Science	5	7	Stories	Stories	Video	1	4	4	5	4
12	5C Faculty Member	Computer Science,Psychology,Neuroscience	4	7	Video	Video	Stories	2	4	4	4	4
12	5C Faculty Member	Computer Science,Psychology,Neuroscience	4	7	Video	Video	Video	1	5	4	4	3
13	5C Student	Computer Science	4	7	Stories	Video	Stories	2	5	4	3	4
13	5C Student	Computer Science	4	7	Stories	Video	Video	2	4	4	3	2
14	5C Student	Science Management	5	7	Video	Stories	Stories	2	4	4	4	5
14	5C Student	Science Management	5	7	Video	Stories	Video	2	5	5	4	5
15	5C Student	Economics,Film Studies	4	3	Video	Video	Stories	3	4	3	5	5
15	5C Student	Economics,Film Studies	4	3	Video	Video	Video	2	5	5	5	5
16	5C Student	Computer Science,Economics	5	8	Stories	Video	Stories	4	5	4	5	5
16	5C Student	Computer Science,Economics	5	8	Stories	Video	Video	4	5	4	5	4
17	5C Student	Computer Science,PPE	5	2	Stories	Stories	Stories	1	5	5	5	5
17	5C Student	Computer Science,PPE	5	2	Stories	Stories	Video	1	5	4	5	4
18	5C Student	Psychology	5	4	Video	Video	Stories	4	4	3	5	4
18	5C Student	Psychology	5	4	Video	Video	Video	5	5	3	5	5
19	5C Student	Computer Science	5	7	Stories	Stories	Stories	3	4	4	4	4
19	5C Student	Computer Science	5	7	Stories	Stories	Video	3	4	3	4	4
2	5C Student	Economics + Engineering (3-2)	4	7	Stories	Video	Stories	2	5	4	4	5
2	5C Student	Economics + Engineering (3-2)	4	7	Stories	Video	Video	2	5	4	5	4
20	5C Student	PPE	1	2	No preference	Stories	Stories	2	5	4	5	5
20	5C Student	PPE	1	2	No preference	Stories	Video	2	5	4	5	4
21	5C Student	Computer Science,Economics	1	1	Stories	No preference	Stories	1	5	5	5	5
21	5C Student	Computer Science,Economics	1	1	Stories	No preference	Video	4	4	4	4	4
22	5C Student	Chemistry	4	5	Stories	Stories	Stories	2	5	5	5	5
22	5C Student	Chemistry	4	5	Stories	Stories	Video	4	2	3	4	4
3	5C Student	Economics + Engineering (3-2)	4	8	Stories	Video	Stories	3	3	4	4	5
3	5C Student	Economics + Engineering (3-2)	4	8	Stories	Video	Video	3	3	4	5	2
4	5C Student	Mathematics,Computer Science	4	5	Video	Video	Stories	2	4	4	4	3
4	5C Student	Mathematics,Computer Science	4	5	Video	Video	Video	2	4	4	4	4
5	5C Student	Computer Science	5	8	Stories	Video	Stories	2	5	4	5	5
5	5C Student	Computer Science	5	8	Stories	Video	Video	4	2	3	5	4
6	5C Student	Mathematics,Computer Science	4	5	Stories	Video	Stories	2	4	4	3	5
6	5C Student	Mathematics,Computer Science	4	5	Stories	Video	Video	4	4	4	5	3
7	5C Student	Computer Science	4	6	Stories	Video	Stories	3	3	3	4	5
7	5C Student	Computer Science	4	6	Stories	Video	Video	2	5	3	4	2
8	5C Student	Computer Science,Economics,Physics	5	5	Stories	No preference	Stories	1	5	5	5	5
8	5C Student	Computer Science,Economics,Physics	5	5	Stories	No preference	Video	3	2	4	4	3
9	5C Student	PPE,French	3	2	Stories	Stories	Stories	2	5	5	5	4
9	5C Student	PPE,French	3	2	Stories	Stories	Video	4	4	4	5	3

User	attention	attention_comments	learning	learning_comments	final_comments
1	Stories	I was able to read it at my pace and fully understand what was happening. By having to tap the screen it ensured that I was actively involved in the experience.	Stories	The story was easy to understand, interactive and I could go at my own pace. The voice in the video was boring, hard to follow and didn't allow me to grasp what was happening due to the pace he was going at.	NA
2	Stories	I liked the interactive component and I could go at my own pace	Video	Even though I like to go at my own pace while learning new information, I often like to have audio and visual components when learning new material.	NA
3	Stories	Because I had some ability to control the pace of the information, I could skip over parts I was done learning about to focus only on the parts that interested me the most.	Video	Less effort involved in the overall experience, which led to a pleasant experience.	NA
4	Video	The sound mostly. I also found his voice entertaining because it wasn't monotone. What I do like about the stories though is the ability to swipe up in the future and get more information about certain aspects of the snap that someone finds interesting. I also find it easier using the snap version to return to where I was and pick back up where as with the video I'm less likely to do that.	Video	I liked the video more because in this study but I think it is primarily because the content was relation to vision work and thus having videos of the demo's were really helpful. For domain transfer / adaptation it helps to have two examples next to each other and that is more difficult to do with the snap version. (Nature of content is what led to video preference).	I think different content is conducive to working better on the snap version relative to video version. I think experimenting with the types of visualizations and figures that work well with the snap version and centering the medium on those will be helpful.
5	Stories	I was more engaged because I had to keep clicking through. It was harder to zone out. Also there was less information so easier to absorb.	Video	Probably the stories because they were more engaging and gave just the key take aways. However, at times the text/description took a while to load which was frustrating because it took longer.	NA
6	Stories	I had to actively participate in the story format by tapping, which allowed me to choose the content that I wanted to focus on and skip over the content that I deemed unnecessary.	Video	It was easier to learn from the video since you do not have to read; you are essentially being taught.	NA
7	Stories	Stories allow you to read at your own pace whereas videos you have to watch the whole way through.	Video	Having a narrator to explain things added more value to the visuals and it added character.	NA
8	Stories	I had more control and was able to absorb the information one fact at a time. When learning through video, if I miss a concept (by a lack of congruency in audio and video for example) then I have to rewind without precision leading me to tediously relearn or spend needless time.	No preference	I like both for different occasions. If I want to effectively focus then stories are preferential and therefore more pleasant. But if I don't want to physically engage and simply get the gist (for review or relaxed learning), then I prefer video.	Stories keep me engaged and allow me to digest concepts with ease. It's harder for me to get engaged with video and it is more frustrating to pause and rewind.
9	Stories	In the videos I was often overwhelmed by needing to both read the text shown in the videos while also listening to the speaker, and so I was not able to grasp as much information or at least was	Stories	While there wasn't a voice speaking to me, with tone intonations which are pleasant to listen to, in the stories, I thought the stories were more	NA

		not able to grasp is as deeply as I was able to in the stories. I also take a while to process things and I liked the ability to move at my own pace with the stories.		pleasant because of how they displayed both text and images in a structured way that was pleasant to follow visually.	
10	Stories	The stories allowed me to go at my own pace which reduced distraction.	Stories	Despite lacking the in-frame experiment footage provided in the video version, the stories were more pleasant to learn from because I did not have to listen and read at the same time. This allowed for stress-free learning and increased retention.	Though I enjoyed the stories version over the video version, I was not particularly impressed by the stories. They included surface level content of academic papers, despite having the freedom to include more sophisticated scientific details via the self-selecting pace aspect. Additionally, I think they could have entirely trumped the video versions by using embedded video clips rather than forcing me to click / swipe to see video on an alternate interface. In conclusion, I feel that the stories were superior learning tools to the videos, but they did not reach the level of specificity I expect out of a scientific paper briefing. But the stories concept definitely has a lot of potential as a pedagogical tool :)
11	Stories	I prefer taking my time when reading through information so it was nice to move on to the next section with a tap at my own pace.	Stories	Again, I'm a slow absorber of information so the stories were better to go at my own pace. I would have liked the stories to include some audio.	As mentioned above, I am both a visual and auditory learner so having some optional audio in the stories would have been nice.
12	Video	voice was engaging, my attention was directed more automatically to where it should be	Video	I generally prefer listening to reading, unless the content is particularly challenging	There are two elements here which are confounded: familiarity with the format & preference for listening vs. reading
13	Stories	Because I needed to tap to learn more, in video format it's easier to let your mind wander	Video	The script from the narrator helped me feel like I understood the topic better, he gave a little extra context to everything	NA
14	Video	Easier to pay attention to the video and listen to the commentary instead of the stories where I had to direct attention to words and then to the video in the background	Stories	Video, I was able to pay attention to what was happening in the video demonstrations while also listening to the commentary	NA
15	Video	It's easier to watch a video and listen to a voice than read	Video	I like having audio and visuals	NA
16	Stories	I could go at my own pace	Video	Having a narrator made me feel like I was being taught. Hearing the positive tone in the narrator's voice made it pleasant and excited	The video often felt too slow. I liked the story because I can go at my own pace and feel more efficient. But the video seemed more natural to learn from.
17	Stories	Easier to go at my own pace	Stories	Again, easier to go at own pace; more easily digestible than a single constant stream	NA
18	Video	i liked the sound	Video	sound	AI is equal parts terrifying and adorable
19	Stories	They are easy to navigate and condensed the information step by step.	Stories	They are easy to navigate and condensed the information step by step.	I think the first video was harder to follow than the second video. Stories were consistently easy to follow.
20	No preference	It was hard to separate the effects of the medium from the effects of viewing the same information again.	Stories	Stories allowed me to go at my own pace.	NA
21	Stories	I think that the story format kept me much more engaged, for I had to actively read and advance the information at my own pace rather than having this done for me.	No preference	I thought that both formats were equally pleasant to learn from. Both the video and story formats both contained the same gifs and images that made the information enjoyable.	NA
22	Stories	I felt more in control of the pace	Stories	again I had more control, feel like I could go about learning at my own pace, my own way.	I enjoyed the stories, more interactive, I could set the own pace and learn the way that accommodated to my learning style. It was like i had all the notes in front of me in a neat format, and could go at my own pace.

C Wireframes

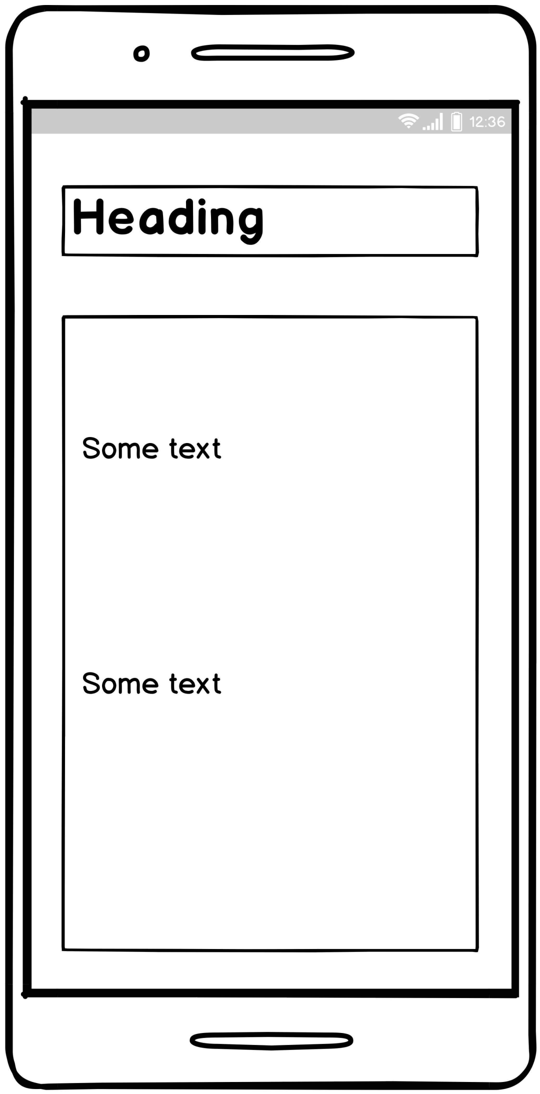




Opening Statement

[1 slide]

- A one-sentence statement explaining why this research is important in clear and simple terms
- Gives high-level context and shows importance
- No technical details
- Ex: "Video is transforming the way we consume scientific ideas."



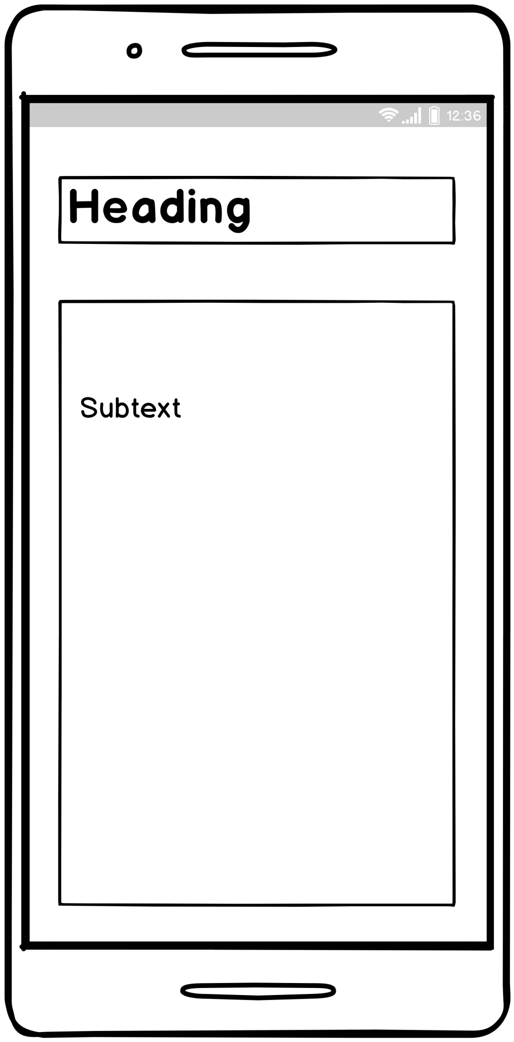
Research Summary

[1 slide]

Useful subsections: Abstract, Introduction

context + very
brief overview of
process
[1-2 sentences]

outcomes
[1 sentence]



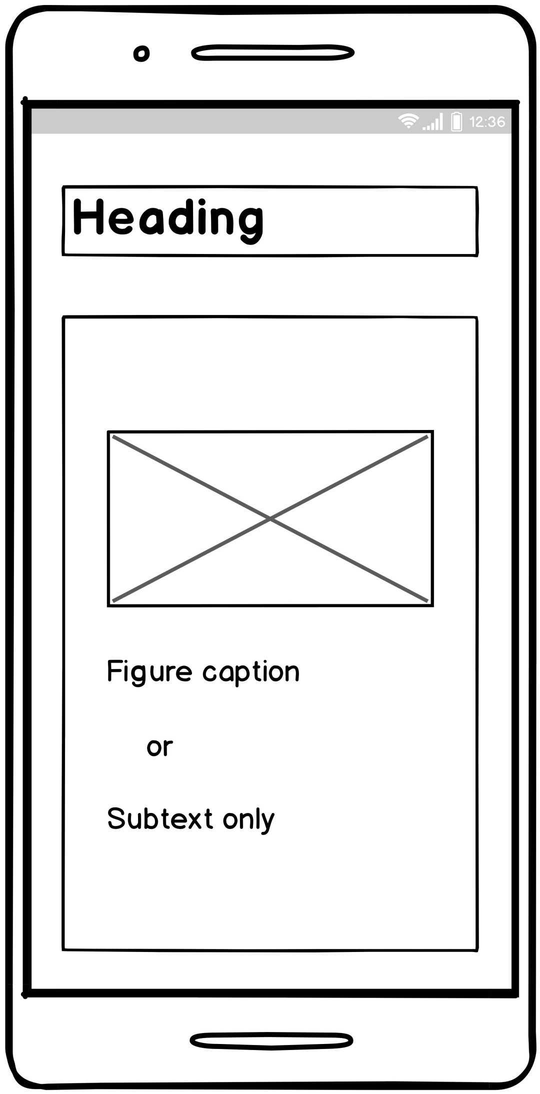
Current State

[1 slide]

Useful subsections: Introduction, Related Work

Out of what existing research and efforts does this research come from? What were the motivations of this research?

+ media if applicable (with caption)



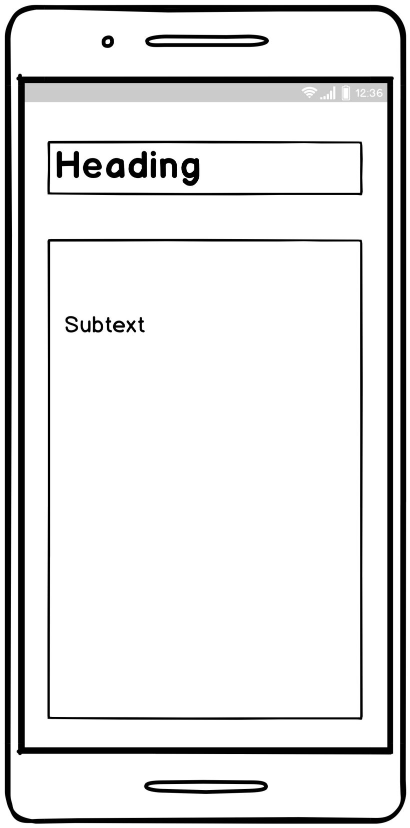
Contributions Summary

[1 slide]

Useful subsections: Abstract, Introduction

Briefly state the contributions and outcomes of what the researchers achieved

+ any relevant media to help with the understanding of the contribution (including a caption to go along with the media)



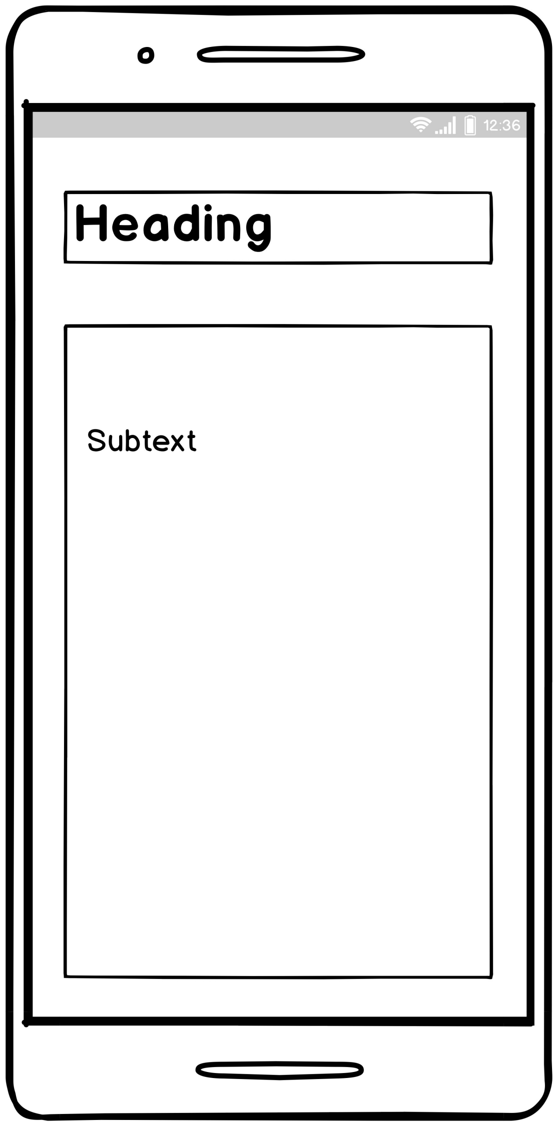
The Process

[1-3 slides]

Useful subsections: Methodology, System Design, Real World Deployment

How did the authors do it?
- Identify and explain the approach they took to answering the research question(s)
- Clarifies the reasons they took this approach

+ any relevant charts or media with captions



Evaluation

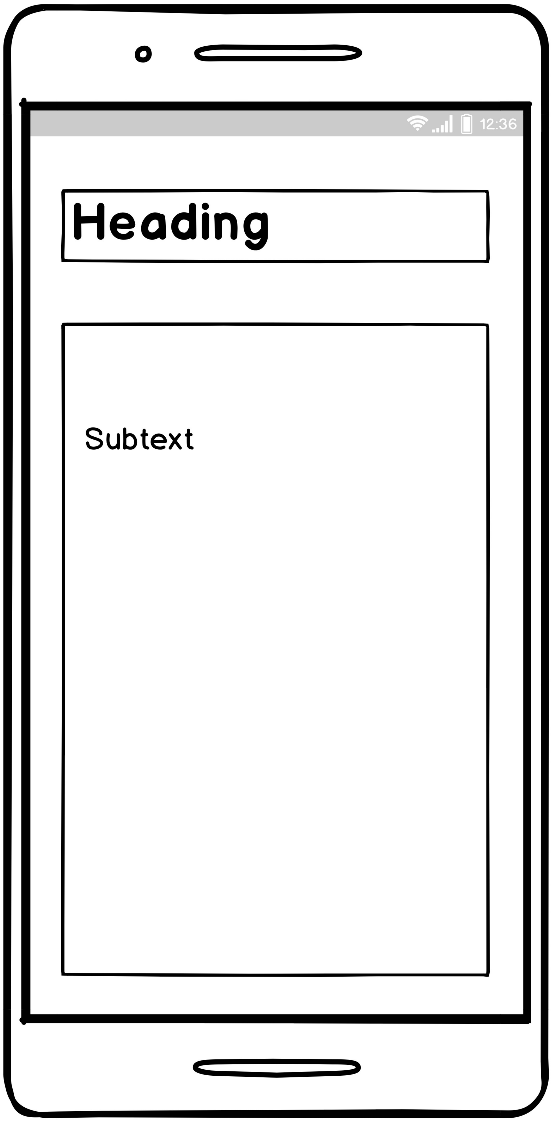
[1 slide]

Useful subsections: Evaluation, Results

Describe how the authors evaluated their claims:

- How did the authors measure the performance of their contribution?
- How does their contribution compare to existing work in the field?

+ any relevant charts or media with captions



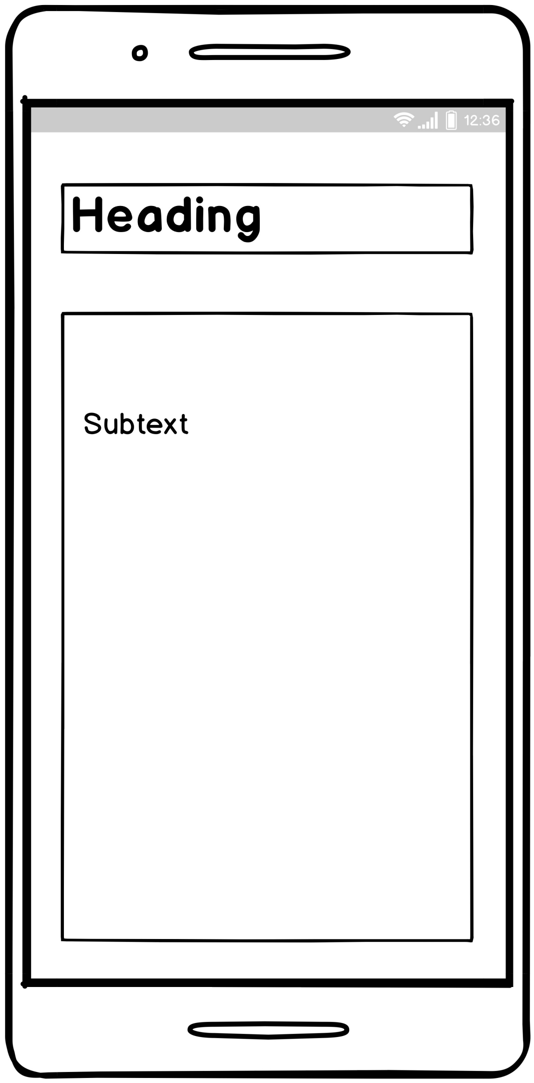
Discovery

[1-2 slides]

Useful subsections: Results, Discussion

Results
- What did the study achieve?
- Chance to go more in depth with the contributions

+ any relevant charts or media with captions



Implications

[1 slide]

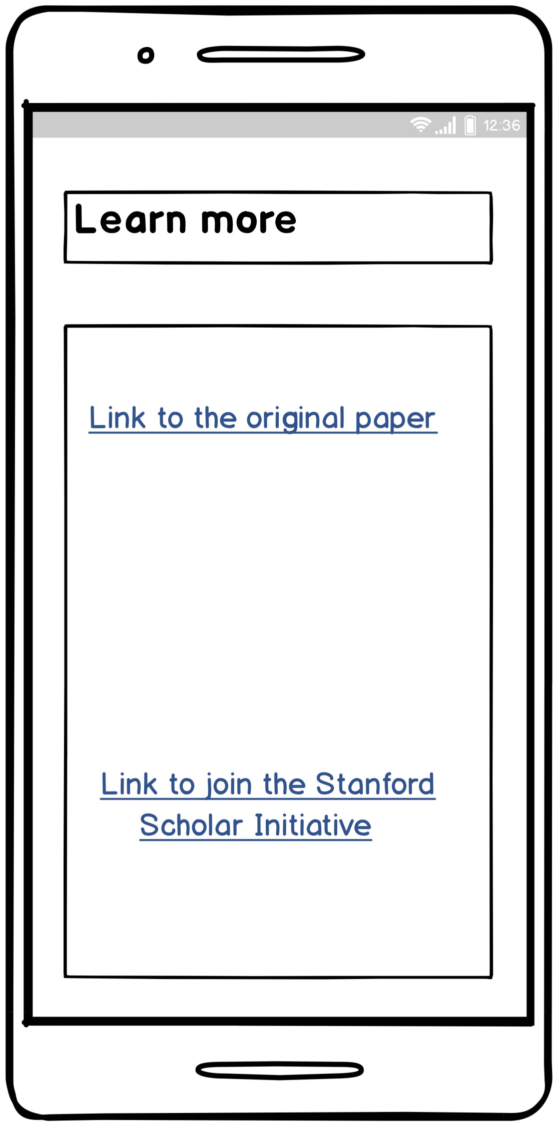
Useful subsections: Discussion, Conclusion

Discussion/Real World Application

- Larger scale contribution to society
- Are there any other applications of the findings?
- What does this mean on a larger scale?

Next Steps

[1 slide]



+ any other relevant references if the reader would like to learn more