Governors State University
OPUS Open Portal to University Scholarship

All Capstone Projects | Student Capstone Projects

Fall 2016

Statistics for Middle and High School Teachers: A Resource for Middle and High School Teachers to Feel Better Prepared To Teach the Common Core State Standards (CCSS) Relating to Statistics

Nanci Kopecky
Governors State University

Follow this and additional works at: http://opus.govst.edu/capstones

 Part of the Statistics and Probability Commons

Recommended Citation
Kopecky, Nanci, "Statistics for Middle and High School Teachers: A Resource for Middle and High School Teachers to Feel Better Prepared To Teach the Common Core State Standards (CCSS) Relating to Statistics" (2016). *All Capstone Projects*. 243.
http://opus.govst.edu/capstones/243

For more information about the academic degree, extended learning, and certificate programs of Governors State University, go to http://www.govst.edu/Academics/Degree_Programs_and_Certifications/

Visit the Governors State Mathematics Department
This Project Summary is brought to you for free and open access by the Student Capstone Projects at OPUS Open Portal to University Scholarship. It has been accepted for inclusion in All Capstone Projects by an authorized administrator of OPUS Open Portal to University Scholarship. For more information, please contact opus@govst.edu.

Statistics for Middle and High School Teachers:
A Resource for Middle and High School Teachers to Feel Better Prepared To Teach the Common Core State Standards (CCSS) Relating to Statistics.

By

Nanci Kopecky

B.A. Mathematics, University of Illinois, Springfield

M.A. Education, Saint Xavier University

Submitted in partial fulfillment of the requirements

For the Degree of Master of Science,

With a Major in Mathematics

Governors State University

University Park, IL. 60484

December 2016

Abstract

The purpose of this project is to create a two-day workshop to better prepare middle and high school teachers to teach probability and statistics as required by the Common Core State Standards (CCSS), which have broadened the mathematics curriculum to include in depth understanding of probability and statistics. Many teachers are not prepared to address probability and statistics concepts. Research has demonstrated a need for greater professional development and resources for teachers in this area. The two-day workshop will allow teachers to review their knowledge and enhance their understanding of statistics by emphasizing student-centered teaching examples. Technology and/or software will be used in connection with the problems. The goal of the workshop is to provide professional development for teachers by helping them to be better qualified and more confident in their ability to teach statistics at the middle and high school level.

Table of Contents:

Table of Tables……4

Table of Appendices……5

Introduction……6

New and Urgent Need for Statistics Education for All……6

Brief and Recent History of Mathematics Practices and Standards……8

Demonstrate the Need for Statistics Education for Teachers……14

Workshop Description……15

Conclusion……19

References……20

Appendices……22

Keywords: professional development, statistics education, teaching high school teachers

Table of Tables:

Table 1. NCTM and CCSSM Standards and Practices.....10

Table 2. NCTM Content Recommendations in Which Secondary Teachers Should Be Proficient.....11

Table 3. Engineering Design Process.....12

Table 4. STEM Practices.....12

Table 5 GAISE Summary.....13

Table 6 Workshop Statistical Concepts.....16

Table 7 NCTM Funneling and Focus Questions.....18

Appendices:

A. Statistics Workshop PowerPoint and Outline Document.....22

Representing Data: How Many Years Does a Penny Stay in Circulation?

B. Worksheet/Activity.....22
C. TInspire Notes.....22

Experimental Design: What is the Readability of the Gettysburg Address?

D. Worksheet/Activity.....22
E. Google Sheets Notes.....22

Normal Distribution

F. Who is a More Elite Athlete? Worksheet.....22
G. What is Normal? Guided Notes and TInspire Notes.....23
H. Geogebra Notes.....23

Linear Regression: Does More Education Translate to Higher Salaries?

I. Worksheet 23
J. Excel Notes.....23

Introduction

The recognition that important decisions are being made in a wide variety of disciplines based on data is causing a reevaluation of the need for statistics education. The education industry is scrambling to create resources to meet the needs of middle and high school teachers to teach statistics. For many professionals, taking mathematics courses in a traditional university setting is not a realistic alternative. Professional development can offer another alternative for teachers to review, learn and update their knowledge of statistics. This paper will examine the need for a basic understanding of statistics and probability for all, give a brief and recent history of the mathematics practices and standards, demonstrate the need for statistics education for teachers, and describe the statistics workshop for middle and high school teachers.

New and Urgent Need for Statistics Education for All

As the importance of data increases, being able to understand statistics and probability is becoming an essential life skill. While mathematicians and statisticians have long understood the importance of statistics and probabilistic thinking, never has the broader society as widely understood the need, including influential political and business leaders. Today, every industry uses data science to make decisions and drive policies. In 2013, Mayer-Schonberger, an Oxford Internet Institute professor, and Cukier, data editor at The Economist, authored a highly influential book, *Big Data: A Revolution that will Transform How We Live, Work, and Think*. New York Times book reviewer, Kakutani (2013) noted:

> Cukier and Mayer-Schonberger argue that big data analytics are revolutionizing the way we see and process the world…they give readers a fascinating-and sometimes alarming-

> survey of big data's growing effect on just about everything: business, government, science and medicine, privacy and even on the way we think (p. C1 in print, para. 8 for online link).

There is a lot of hope and excitement regarding big data; proponents of big data argue that it will make us healthier, safer, smarter and more efficient. There is a growing number of jobs in statistics and data science. Google Chief Economist Hal Varian is widely quoted as saying that "the sexy job in the next 10 years will be statisticians. People think I'm joking" (Davenport & Patil, 2012, para. 25).

The use of statistics and the explosive expansion of data science is exciting, but it is important to recognize its limitations. One danger is data's seemingly unrestrained influence on decision making and policies. Data should not be valued over people. The effects of policies driven by big data are just emerging. Harvard Business Review contributors, McAfee and Brynjolfsson (2012) state, "Big data's power does not erase the need for vision or human insight" (para. 19). In the past, knowledge of a profession and how to use data were separate entities, but now it is a merging skill set. Waller and Fawcett (2013) contend finding the balance of the two is key to moving any given domain forward.

There is promise of making a brighter future with the use of statistics and data, and now, more than ever, there is responsibility to do it in an ethical way. Consumers can be taken advantage of by the barrage of claims from daily headlines. Human rights can be violated, such as informed consent and privacy. In 2014, Facebook studied how positive and negative newsfeeds influenced nearly 700,000 users (Goel, 2014). Checking the "I Agree" box to use software hardly qualifies for the informed part of informed consent; there is no full disclosure. While the Facebook study meets the gold standards for a large, double blind, controlled, and

randomized experiment, it violated the law of informed consent, had the potential to cause harm[1], and attempted to influence what we think.

Collecting and analyzing big data to improve the quality of life and the environment is a benevolent goal. The potential to solve problems assumes that the data collected is accurate, but the pressure to lie or mischaracterize information and events is real. Growing reports indicated that administrators and politicians pressure professionals in the medical, public safety, and education fields to falsify data. In Chicago, officers reported that they were asked to reclassify incident reports to show neighborhoods were getting safer (Bernstein & Jackson, 2014). Educators in Georgia went to jail for changing the answers to standardized tests (Fantz, 2015).

Statistics education for the masses is a key component to ensure individuals are treated with integrity and do not become hardened by today's complex and exponentially fast paced world. Therefore, it is imperative for the public to have some sense of how their individual daily actions, such as use of social media, software, and apps on any electronic device, influence the world around them. The clicks and swipes of our fingertips send data through thin air and fiber optics. It is then measured by computers and used by humans, which, in turn, affects all our senses and what we think. Society now calls for statistics education for all!

Brief and Recent History of Mathematics Practices and Standards

[1] We do not know if anyone was mentally or physically harmed by this study. Is it possible that of the 689,003 participants, one was suicidal or depressed and sensitive to a negative newsfeed? Maybe. In fact, the findings of the study show the mood of the user was affected by the biased newsfeed. What gives Facebook the right to ruin someone's day?

In 2010, President Obama's administration made public the Common Core State Standards (CCSS), which were intended to provide a globally competitive and uniform curriculum across the country. The genesis of CCSS was the push for national standards from the National Governors Association in the early 1990s. At that time, there were significant concerns about American students' preparation to enter college or the workforce (Wiki, 2016). Corporations, universities and professional organizations played a role in developing the standards and practices over the years.

The mathematical standards and practices included in the Common Core State Standards for Mathematics (CCSSM) are primarily derived from the National Council of Teachers of Mathematics (NCTM) guidelines. Leinwand, Brahier, and Huinker (2014), professors on a writing team for NCTM, show the historical progress of the improved math curriculum:

> In 1989, the NCTM launched the standards-based education movement in North America with the release of Curriculum and Evaluation Standards for School Math, an unprecedented initiative to promote systemic improvement in math education. Now, twenty-five years later, with wide spread adoption of college and career readiness standards, including adoption in the U.S. of the CCSSM by forty-five states… provides an opportunity to reenergize and focus our commitment to significant improvement in mathematics education (p. 1).

NCTM structured its approach to mathematics education in four categories: standards, practices, principals, and processes. Standards detail the content to be studied. Practices describe how the content should be taught. Principles[2] are overarching themes to guide decision making and mathematical processes[3] describe the applications of knowledge. CCSSM

[2] NCTM principles are equity, curriculum, teaching, learning, assessment, and technology.

[3] NCTM processes are problem solving, reasoning and proof, communication, connections and representation.

offered a more condensed version that is set out in a set of detailed standards and practices. Table 1 compares NCTM and CCSSM standards and practices (relating to middle and high school statistics and probability).

Table 1

NCTM and CCSSM Standards and Practices

NCTM Standards	CCSSM Standards	NCTM and CCSSM Practices
1. Formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics with one population.	Develop understanding of statistical variability.	Make sense of problems and persevere in solving them.
2. Select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.	Summarize and describe distributions.	Reason abstractly and quantitatively.
3. Understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each.	Use random sampling to draw inferences about a population. Draw informal comparative inferences about two populations.	Construct viable arguments and critique the reasoning of others.
4. Understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each. *	Investigate chance processes and develop, use, and evaluate probability models. Investigate patterns of association in bivariate data.	Model with math.
5. Know the characteristics of well-designed studies, including the role of randomization in surveys and experiments. *	Interpreting Categorical and Quantitative Data*	Use appropriate tools strategically.
6. Understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable. *	Making Inferences and Justifying Conclusions*	Attend to precision.
7. Understand histograms, parallel box plots, and scatterplots and use them to display data. *	Conditional Probability and the Rules of Probability*	Look for and make use of structure.
8. Compute basic statistics and understand the distinction between a statistic and a parameter. *	Using Probability to Make Decisions*	Look for and express regularity in repeated reasoning.

*High school standards

Table 2 lists the statistics and probability topics recommended by NCTM in order for secondary teachers to be proficient.

Table 2.

NCTM Content Recommendations in which Secondary Teachers Should Be Proficient

1. Statistical variability and its sources and the role of randomness in statistical inference
2. Creation and implementation of surveys and investigations using sampling methods and statistical designs, statistical inference (estimation of population parameters and hypotheses testing), justification of conclusions, and generalization of results
3. Univariate and bivariate data distributions for categorical data and for discrete and continuous random variables, including representations, construction and interpretation of graphical displays (e.g. box plots, histograms, cumulative frequency plots, scatter plots, summary measures, and comparisons of distributions
4. Empirical and theoretical probability (discrete, continuous, and conditional) for both simple and compound events
5. Random(chance) phenomena, simulations, and probability distributions and their application as models of real phenomena and to decision making
6. Historical development and perspectives of statistics and probability including contributions of significant figures and diverse cultures

In addition to CCCSS, another main driver in the national discussion on mathematics has been the broader discussion regarding Science, Technology, Engineering, and Mathematics (STEM) programs. The National Science Foundation (NSF) began funding projects based on the NCTM standards in the early 2000s, which contributed to the articulation of STEM-based curriculum. Elaine J. Hom (2014), LiveScience.com contributor, described STEM as “a curriculum based on the idea of educating students in four specific disciplines — science, technology, engineering and mathematics — in an interdisciplinary and applied approach” (para. 1). STEM solves real-life problems and encourages open ended exploring in a collaborative team effort. It incorporated an Engineering Design Process (EDP), which like the NCTM and CCSSM mathematical practices, begins by defining a problem, develops and tests a possible

model/solution, and then redesigns or reflects on the model. It is cyclic and iterative; failure is part of the process. Table 3 contains an image of the EDP developed by middleweb.com.

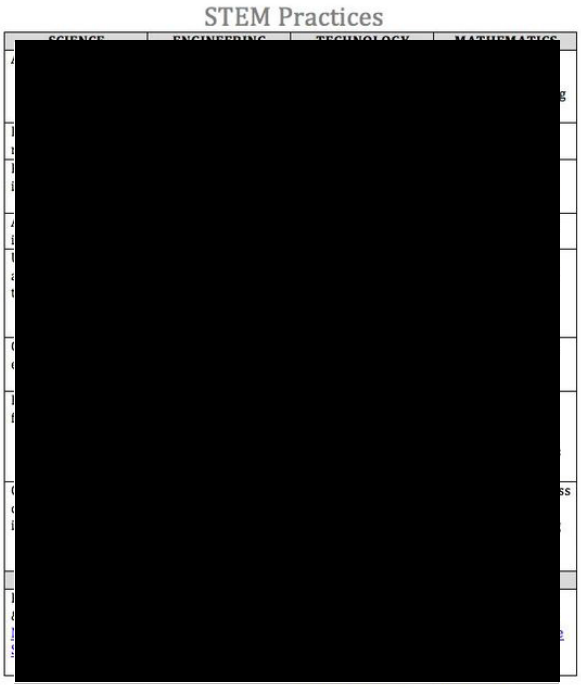
Table 3

Engineer Design Process

Table 4 contains a chart created by a science educator and blogger, Robbins (2013), summarizing how STEM is applied to different disciplines.

Table 4

STEM Practices



In addition to CCSS and STEM proponents, another important voice regarding the need for improved statistics education is the American Statisticians Association (ASA). ASA has prepared Guidelines for Assessment and Instruction in Statistics Education (GAISE). The GAISE framework includes different terminology – identifying recommendations and goals for universal statistics education rather than standards, practices, and principles and processes). However, their Four Step Statistical Process reflects the scientific method, EDP, and mathematical processes discussed above in connection with other standards. Table 5 presents a summary of GAISE.

Table 5.

GAISE Summary

Recommendations	Goals, abbreviated

1. Teach statistical thinking.
   A. 4 Step Statistical Process
      i. Formulate a question that can answered by data
      ii. Design and implement a plan to collect appropriate data
      iii. Analyze the collected data by graphical and numerical methods
      iv. Interpret the analysis in the context of the original question
2. Focus on conceptual understanding.
3. Integrate real data with a context and a purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

1. Be critical consumers
2. Know when statistics is useful for the investigative process
3. Represent data and interpret graphs and numerical summaries
4. Understand variability
5. Understand randomness
6. Gain experience with statistical models
7. Use statistical inference
8. Draw conclusions using statistical software
9. Aware of ethical issues regarding statistical practice

As suggested by the discussion above, there are many professional organizations and government panels having important discussions on preparing students for 21st century mathematics. Although they share a common desire to promote mathematics education, each group arranges the mathematical content and teaching methods in different formats, which can make the discussion somewhat difficult to navigate. Fortunately, there are common themes through each of the suggested frameworks. There appears to be broad consensus on appropriate content, the need to develop students' intuition about the content and to offer opportunities to explore problem solving and analysis.

Demonstrate the Need for Statistics Education for Teachers

Many middle and secondary mathematics teachers were overwhelmed with the new statistics standards from the common core. Many teachers have had limited, or no, exposure to statistics courses in college. For those who have some exposure, it has often been a long time since they had have taken a class. Gould and Peck (2004), statistics professors at

University of California, Los Angeles and California Polytechnic respectively, demonstrated the need for statistics education for mathematics teachers and researched ways to increase their statistical content knowledge. They referred to secondary teachers' "lack of statistical literacy," and stated that the "the need for qualified teachers is growing." Cannon (2016), statistics professor at Cornell College in Iowa states, "One of the positive aspects of the Common Core curriculum is its focus on data and data analysis. However, many of our teachers have limited training in statistics. They are being asked to teach information that they have not been exposed to themselves" (para. 5).

The ASA released Mathematical Education of Teachers (MET) II in 2012, an updated report of MET I (2001), which detailed what mathematics teachers need to know and how to be prepared. The MET II included the 2010 common core state standards and emphasized the need for statistics education. The Statistical Education of Teachers (SET) followed with more details regarding standards, pedagogy, examples and resources geared specifically for statistics (Franklin, et. Al, 2015).

Both reports, the MET II and SET, demonstrated the need for teacher training as well as offering recommendations for addressing the problem. The reports emphasized a collaborative effort between professional statisticians and mathematicians in education to bring middle and secondary mathematics teachers skills up to par.

> [The MET II] urges greater involvement of mathematicians and statisticians in teacher education so that the nation's mathematics teachers have the knowledge, skills, and dispositions needed to provide students with a mathematics education that ensures high school graduates are college- and career-ready as envisioned by the Common Core State Standards (p. xi).

The MET II contended that teachers need expert skill in understanding statistics. Students need to go beyond the computation methodology and learn the methodology of structural arguments in composing a survey or study. Teachers need to be able to do more than identify computational errors.

> Teachers need the ability to find flaws in students' arguments, and to help their students understand the nature of the errors. Teachers need to know the structures that occur in school mathematics, and to help students perceive them (p. 2).

Workshop Description

The goal of the workshop is to instill confidence and prepare middle and high school teachers to better address the challenge of teaching statistics. Teachers will understand the policies driving the curriculum changes and will be exposed to current information from professional organizations driving curriculum changes. The workshop will provide opportunities for teachers to review and learn statistics through student-centered examples. Software and technology that is relevant to teaching in the classroom and utilized in the workforce will be incorporated into each problem. The workshop is outlined in a PowerPoint (see Appendix A). Table 6 shows the statistical concepts, mathematics problem and technology that are integrated together for this workshop.

Table 6 Workshop Statistical Concepts

Statistical Concepts	Math Problem	Technology/ Software

Representing Data Graphically and Numerically	How Many Years Does a Penny Stay in Circulation? See Appendix B	TInspire See Appendix C
Experimental Design	What is the Readability of the Gettysburg Address? See Appendix D	Google Sheets See Appendix E
Normal Distribution	Who is a More Elite Athlete? See Appendix F What is Standard Normal with TInspire? See Appendix G	Geogebra See Appendix H
Linear Regression	Does More Education Translate to Higher Salaries? See Appendix I	Excel Worksheet See Appendix J

The student-centered examples are based on the concept of Student-Centered Learning (SCL) which supports the STEM model, as well as incorporating traditional teaching styles and philosophies.

> Educators will need to make significant instructional shifts to help students reach standards that emphasize not only application of mathematical procedures, but also deep understanding, problem solving, critical thinking, and communication… student-centered learning consists of an array of complementary approaches to teaching and learning that draws from multiple theories, disciplines, and trends in the field of education (Walters, et al., 2014, p. 5).

In SCL, students use their prior knowledge and reasoning skills to answer the how and why questions. The study by Walters et. al, found that students exposed to SCL reported being more engaged and interested in the subject matter. They also had higher PISA assessment scores (2014).

Walters et. al provides a transcript between a teacher and student that demonstrates student-centered questioning. In the transcript, the student thinks aloud answering her own

questions and perseveres through the problem with minimum guidance by the teacher. The teacher's comments prompt the student to continue to answers their own questions through logic and visually picturing the problem.

NCTM also encourages teachers to use instruction that emphasizes students intuition in identify problems, thinking of ways to solve them, then modeling solutions. NCTM refers the type of questioning that supports this type of instruction as "focusing," as opposed to "funneling" (2014). Both questioning methods are valuable; focusing questions are more conducive to the STEM model and CCSS mathematical practices. In funneling questioning, the teacher begins with probing questions, followed by higher level questions. Student answers can be computational or superficial because there is not enough time for responses, (NCTM, 2014). Focusing questioning students are pushed to clarify there ideas and identify where the mathematics is in the problem. Table 7 shows the different types of questioning.

Table 7 NCTM Funneling and Focus Questions.

The workshop will explore the examples intuitively, as well as through the use of a variety of software and technology. The examples and steps on how to use the technology and software are in Appendices B through J. It was considered focusing on one tool, such as, the TInspire®, rather than a little exposure on several software applications. However, it is important for educators to see the similarity of many of these programs so that exploring software will not be overwhelming. This will allow teachers to work more with 21st century mathematics.

Conclusion

There is much for today's middle and high school teachers to navigate through including keeping up to date on curriculum, mastering new software and technology, and implementing new teaching methods. This workshop was designed to help prepare middle and high school teachers to teach statistics. By reviewing and teaching key statistical concepts with current trends in teaching methods, and the use of popular software and technology, teachers will be better equipped for today's data-driven classroom. This paper examined the new and urgent need for statistics education for all, gave a brief and recent history of the mathematics practices and standards, demonstrated the need for statistics education for all teachers, and described a statistics workshop for middle and high school teachers.

References

Achieve. (2004) Ready or Not: Creating a High School Diploma That Counts. http://www.achieve.org/ReadyorNot

Bernstein, David & Jackson, Noah. (2015, April 7) The Truth about Chicago's Crime Rates. *The Chicago Magazine.*

Beckmann, S., Chazan, D., Cuoco, A. Fennell, F., Findell, B., Kessel, C., King, K., Lewis, W.J., McCallum, W., Papick, I., Reys, B., Rosier, R., Schaeffer, R., Spangler, D. & Tucker, A. (2012). The Mathematical Education of Teachers II (MET-II). *Conference Board of Mathematical Sciences: Issues in Mathematics Education.* (17). http://cbmsweb.org/MET2/met2.pdf

Connor, Ann. (2016, October 21). *Fund the Future: Support Statistics.* American Statistical Association. https://www.amstat.org/ASA/Giving/Ann-Cannon.aspx

Davenport, T.H. & Patil, D.J. (2012, October) Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. (Paragraph 25 for online link).

Goel, Vindu. (2014, June 29). Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. *New York Times*

Hom, Elaine J. (2014, Feb 11) What is Stem Education? http://www.livescience.com/43296-what-is-stem-education.html

Fantz, Ashley. (2014, April 14). Prison time for some Atlanta school educators in cheating scandal. CNN.

Franklin, Christine A., Bergagliottie, Anna E., Case, Catherine A., Kader, Gary D., Schaeffer, Richard L., Spangler, Denise A. (2015) The Statistical Education of Teachers (SET). *American Statistical Association (ASA)* website. http://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

Gould, Robert & Peck, Roxy. (2014). Preparing Secondary Mathematics Educators to Teach Statistics. *Curricular Development Statistics Education*. Sweden.

Kakutani, Michiko. (2013, June 10). Watched by the Web: Surveillance Is Reborn. *New York Times* (page C1 in print or paragraph 8 online)

Leinwand, Steven, Brahier, Daniel J., Huinker, DeAnn (2014). *Principles to Actions: Ensuring Mathematical Success For All.* Restin, VA: National Council of Teachers of Mathematics.

McAfee, Andrew & Brynjolfsson, Erik (October 2012) Big Data: The Management Revolution. Harvard Business Review.

MiddleWeb. (2015) Engineering Design Process image http://www.middleweb.com/23308/teaching-stem-summer-planning-pays-off/

Robbins, Kirk. (2013, June 11). STEM Practices (included chart). *Science for All* website. https://teachscience4all.org/2013/06/11/stem-practices/

Waller, Matthew A. & Fawcett, Stanley E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design*, Journal of Business Logistics*, *34(2)*, 77-84.

Walters, K., Smith, T.M., Leinwand, S., Surr, W., Stein, A. & Bailey, P. (November 2014). An Up-Close Look at Student-Centered Mathematics Teaching: A Study of Highly Regarded High School Teachers and Their Students. NMEfoundation.org

Wiki (2016). Common Core State Standards Initiative. https://en.wikipedia.org/wiki/Common_Core_State_Standards_Initiative

Appendix B

Names _____

Representing Data

Materials: 20 pennies per group, Double sided tape, Poster Board

Question: How many years does a penny stay in circulation?

Part I. Collect and Represent the Data in at least two different formats.

Q1. Describe some of the pros and cons of the different ways to display data.

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Q2. What is your preferred way to display data? Explain.

Part II. After you collect your data, place your penny in the class Penny Poster in the front of the classroom to create a dot plot.

Part III. Analyze the Data, SOCS

SOCS	**Your Sample, n=20**	**Class Sample, n=**
Shape		
Outliers		
Center		
Spread		

Q3. State at least three things do you notice or wonder about the age of the pennies.

Q4. Compare and contrast your sample and the class sample.

Part IV. Conclusions

Q4. What are some conclusions you can draw about the age of the pennies?

Q5. What are the limitations of this study?

Q6. Can you generalize the above conclusions to the entire penny population? Justify your answer.

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Appendix B

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Appendix C

Technology Notes for *TI nspire*

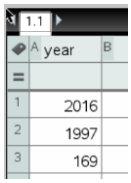
(How Many Years doe a Penny in Circulation?)

Question: How many years does a penny stay in circulation?

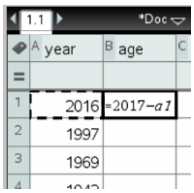
1. Compute Descriptive Statistics

Select 1. New Document, 2. List and Spreadsheets

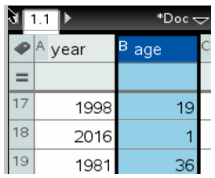
Enter the data Column A (Option: Name Column A "year" by selecting the top cell)



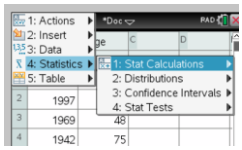
Enter in cell B1: $= 2017 - a1$ (Instead of typing a1, select the a1 cell), press Enter

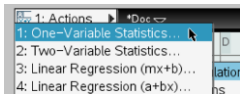


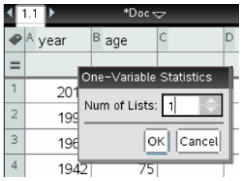
Select B1 to highlight**, hover arrow** over bottom right corner of cell until **plus sign appears, click, and drag down** until cell b20. (2017-a# will repeat). Click another part of the page to remove highlight from column B.



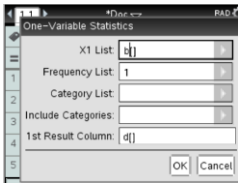
Select Menu, 4. Statistics, 1. Stat Calculations, 1. One Variable Statistics, OK



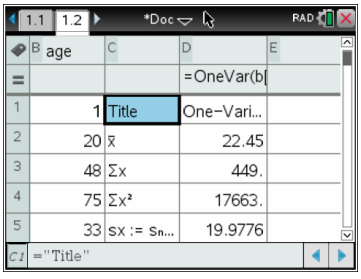




Enter in X1 List: $b[\,]$, **Select** OK.



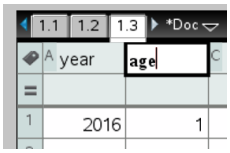
Results:



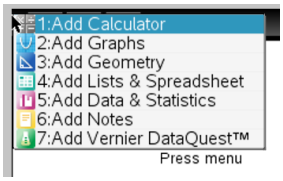
Note: For more advanced use, become familiar with the commands, $= OneVar(b[],1)$

For Results in more reader friendly format:

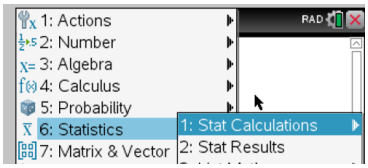
Select "age" cell, or cell at the top of column B, **Press** ctrl, sto→, and **type** age

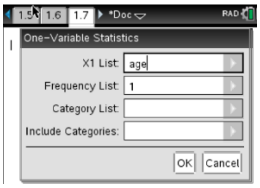


Press cntl, +page, 1: Add Calculator.

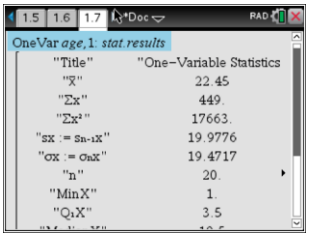


Press Menu, 6: Statistics, 1: Stat Calculations, 1: One variable Statistics, **Select** OK, **Enter in X1 List:** age **Select** OK,



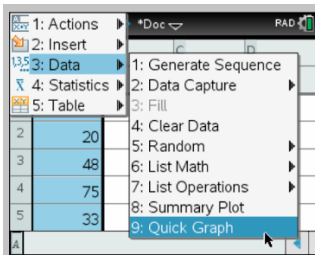


Results:

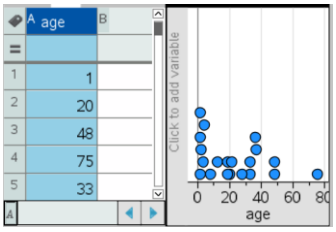


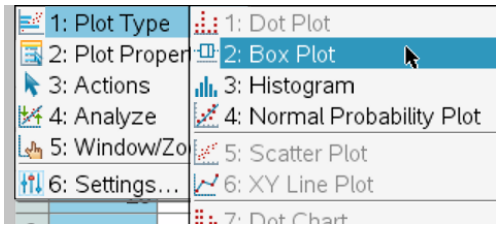
2. Represent the Data

Select and Highlight "age" column by clicking on the right line of the cell, **Select** Menu, 3: Data, 9: Quick Graph



Results: Change the graph type. **Click on graph, Select** Menu, 1: Plot Type, and then 2: Box Plot or 3: Histogram





To Undo Split Screen:

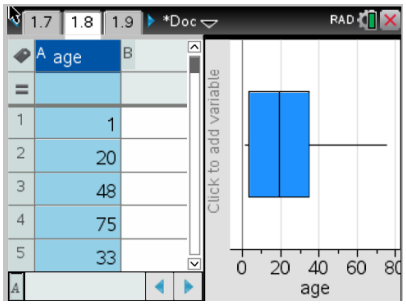
Doc,

5: Page Layout,

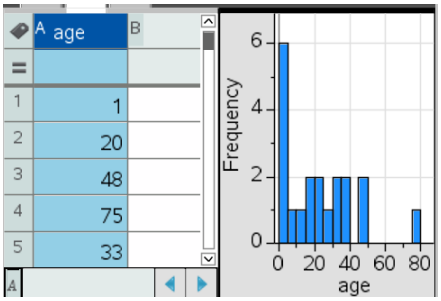
8: Ungroup

More Results in Split Screen:

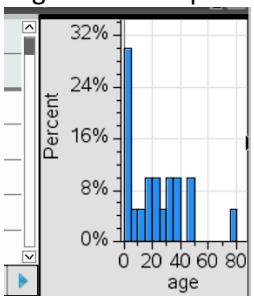
Box Plot



Histogram, Frequency Chart

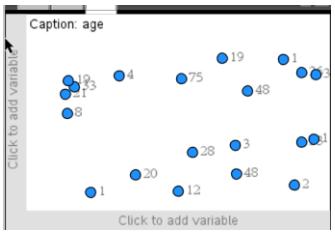


Relative Frequency Chart (change 4:Scale to percent)

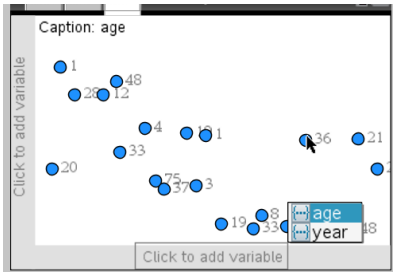


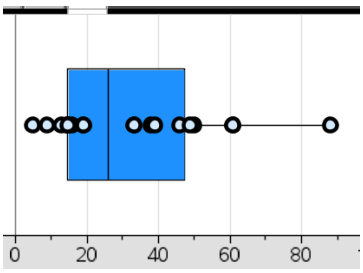
Results on a New Page:

Press ctrl, +page, 5: Data & Statistics, Enter.



Click to add variable, Select age, Enter. For more graph types, **Press** Menu, 1: Plot Type, 2: Box Plot or 3: Histogram.





Handy Tips:	
Ctrl Z	Undo
Ctrl Menu	Right Click
Ctrl Arrow	Move between tabs
Doc, 5: 5:	Delete a tab
Doc, 5: 8:	Undo Split Screen

Appendix D

Names _____

Sampling

Question: What is the readability of the Gettysburg Address?

Directions: Average word length is one of several measurements to help determine the readability of a passage. Find the average work length of the Gettysburg Address, a historic and world famous speech by Abraham Lincoln given at the Cemetery for Gettysburg in an attempt to give meaning to the events that took place one of which being over 50,000 causalities.

The Gettysburg Address

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate—we can not consecrate—we can not hallow—this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

Part I.

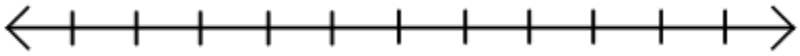
1. Select a random sample of 5 random words from The Gettysburg Address and circle them. Represent the data in a table (A) and dot plot (B).

Table (A)

n = 5	1	2	3	4	5
Word					
# of letters					

Appendix D

Box Whisker (B), n =5



(Based on the graph, estimate the mean.)

Indicate the following:

Observational units:

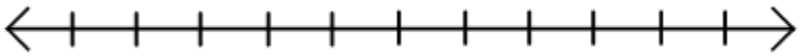
Variable: Continuous or Discreet

Type of Variable: categorical/numerical Sample

Your sample mean (average): $\bar{x} =$

2. At the front board, make a class dot plot (C), put your sample average on a dot plot with the rest of the class

Dot Plot (C)



(Based on the class sample dot plot, estimate the mean)

Now, compute the class sample mean (use *Google Sheets*), $\bar{x} =$

3. The population average (mean), μ is _____. Plot the population average on your Dot Plots (B) and (C). Note the population, N, of the Gettysburg Address is 268 words.

4. What are your observations about the sample means compared with the population mean?

   a. Were the sample means was accurate?

   b. Was one of the sample means more accurate than the other? If so, why do you think?

c. List some ways to improve your sampling so the sample mean(s) will be more accurate.

The **central limit theorem** states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough, $as\ n \to \infty$.

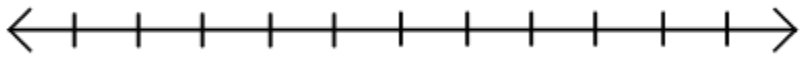
Part 2.

5. Use a computer to generate 5 random numbers between 1 and 268. It is ok to have a repeating number. Use The Gettysburg Address numbered word chart to find the corresponding word to the computer generated random number.

n = 5	1	2	3	4	5
Generated Random Number					
Word					
# of letters					

Your Sample Average (mean), $\bar{x}$:

6. Combine your sample average, $\bar{x}$, with the class on dot plot on the front board. (D).



7. By changing our sampling method change the results of our class sample mean? What conclusions can you make about sampling?

Sampling Frame

1	Four	55	We	109	cannot	163	for	217	they
2	score	56	are	110	dedicate	164	us	218	gave
3	and	57	met	111	we	165	the	219	the
4	seven	58	on	112	cannot	166	living	220	last
5	years	59	a	113	consecrate	167	rather	221	full
6	ago	60	great	114	we	168	to	222	measure
7	our	61	battlefield	115	cannot	169	be	223	of
8	fathers	62	of	116	hallow	170	dedicated	224	devotion
9	brought	63	that	117	this	171	hear	225	that
10	forth	64	war	118	ground	172	to	226	we
11	upon	65	We	119	The	173	the	227	here
12	this	66	have	120	brave	174	unfinished	228	highly
13	continent	67	come	121	men	175	work	229	resolve
14	a	68	to	122	living	176	which	230	that
15	new	69	dedicate	123	and	177	they	231	these
16	nation	70	a	124	dead	178	who	232	dead
17	conceived	71	portion	125	who	179	fought	233	shall
18	in	72	of	126	struggled	180	here	234	not
19	liberty	73	that	127	here	181	have	235	have
20	and	74	field	128	have	182	thus	236	died
21	dedicated	75	as	129	consecrated	183	far	237	in
22	to	76	a	130	it	184	so	238	vain
23	the	77	final	131	far	185	nobly	239	that
24	proposition	78	resting	132	above	186	advanced	240	this
25	that	79	place	133	our	187	It	241	nation
26	all	80	for	134	poor	188	is	242	under
27	men	81	those	135	power	189	rather	243	God
28	are	82	who	136	to	190	for	244	shall
29	created	83	here	137	add	191	us	245	have
30	equal	84	gave	138	or	192	to	246	a
31	Now	85	their	139	detract	193	be	247	new
32	we	86	lives	140	The	194	here	248	birth
33	are	87	that	141	world	195	dedicated	249	of
34	engaged	88	that	142	will	196	to	250	freedom
35	in	89	nation	143	little	197	the	251	and
36	a	90	might	144	note	198	great	252	that
37	great	91	live	145	nor	199	task	253	government
38	civil	92	It	146	long	200	remaining	254	of
39	war	93	is	147	remember	201	before	255	the
40	testing	94	altogether	148	what	202	us	256	people
41	whether	95	fitting	149	we	203	that	257	by
42	that	96	and	150	say	204	from	258	the
43	nation	97	proper	151	here	205	these	259	people
44	or	98	that	152	but	206	honored	260	for
45	any	99	we	153	it	207	dead	261	the
46	nation	100	should	154	can	208	we	262	people
47	so	101	do	155	never	209	take	263	shall
48	conceived	102	this	156	forget	210	increased	264	not
49	and	103	But	157	what	211	devotion	265	perish
50	so	104	in	158	they	212	to	266	from
51	dedicated	105	a	159	did	213	that	267	the
52	can	106	larger	160	here	214	cause	268	earth
53	long	107	sense	161	It	215	for		
54	endure	108	we	162	is	216	which		

Statistics for Middle and High School Teachers

A Resource for Middle and High School Teachers to Feel Better Prepared to Teach the Common Core State Standards (CCSS) relating to Statistics.

Welcome!

- Pick up a binder and Tinspire calculator
- Fill out name tags
- Find our online classroom at Schoology.com
  - Class code:
  - Begin Survey Monkey
- Introductions
  - Name
  - Education experience
  - A hobby or interest

Workshop Goal: Prepare Middle and High School Teachers to Teach Statistics

- Review and Learn Statistics through Student-Centered Problems
- Represent and analyze data with the software and technology
  - TInspire
  - Excel
  - Google Sheets
  - Khanacademy.com
  - Wolfram Mathematica
  - Desmos.com
  - Geogebra.com

Workshop Outline

- Mathematics Curriculum and Practices Today
- Representing Data
- Experimental Design
- Normal Distribution
- Linear Regression

Mathematics Curriculum and Practices Today

- (CCSS) Common Core State Standards
- (NCTM) National Council of Teaches of Mathematics
- (STEM) Science Technology Engineering and Mathematics
  - Concept began in meetings sponsored by the National Science Foundation (NSA)
- (GAISE) Guidelines for Assessment and Instruction in Statistics Education
  - By American Statisticians Association (ASA)
- Connection
  - The CCSS derived from the NCTM standards which were developed in 1989 and updated in 2000. The NCTM standards were praised in the education and science industries. The National Science Foundations funded projects based on these standards. STEM came out of meetings held by the National Science Foundation and started showing up in the early 2000's.

CCSSM-Probability and Statistics

Standards

- Develop understanding of statistical variability.
- Summarize and describe distributions.
- Use random sampling to draw inferences about a population.
- Draw informal comparative inferences about two populations.
- Investigate chance processes and develop, use, and evaluate probability models.
- Investigate patterns of association in bivariate data.
- Interpreting Categorical and Quantitative Data*
- Making Inferences and Justifying Conclusions*
- Conditional Probability and the Rules of Probability*
- Using Probability to Make Decisions*

8 Mathematical Practices

- Make sense of problems and persevere in solving them.
- Reason abstractly and quantitatively.
- Construct viable arguments and critique the reasoning of others.
- Model with mathematics.
- Use appropriate tools strategically.
- Attend to precision.
- Look for and make use of structure.
- Look for and express regularity in repeated reasoning.

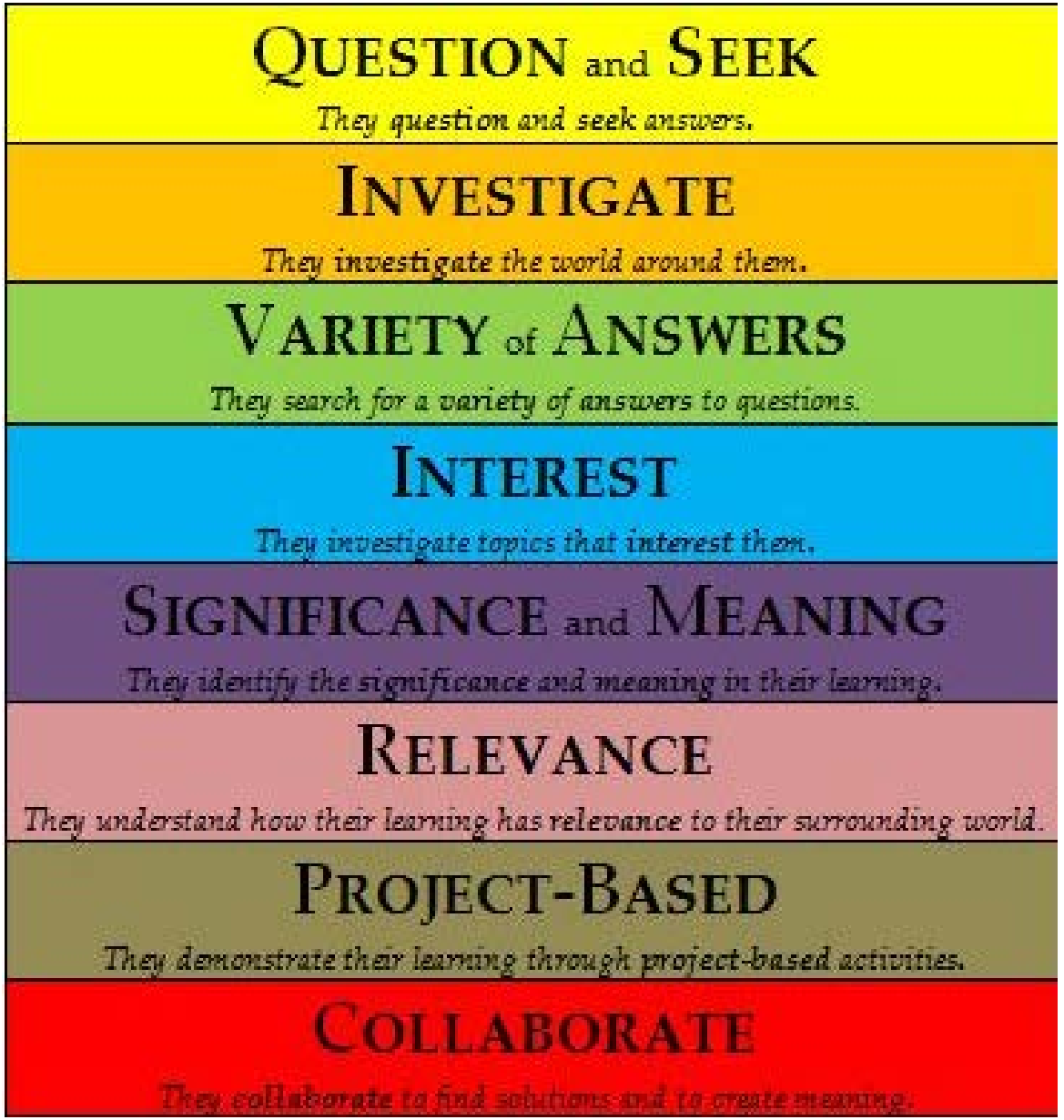
NCTM – Data Analysis and Probability

- **Summary of Standards and Principles**
  - Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them
  - Select and use appropriate statistical methods to analyze data
  - Develop and evaluate inferences and predictions that are based on data
  - Understand and apply basic concepts of probability

NCTM-Detailed Data Analysis and Probability Standards

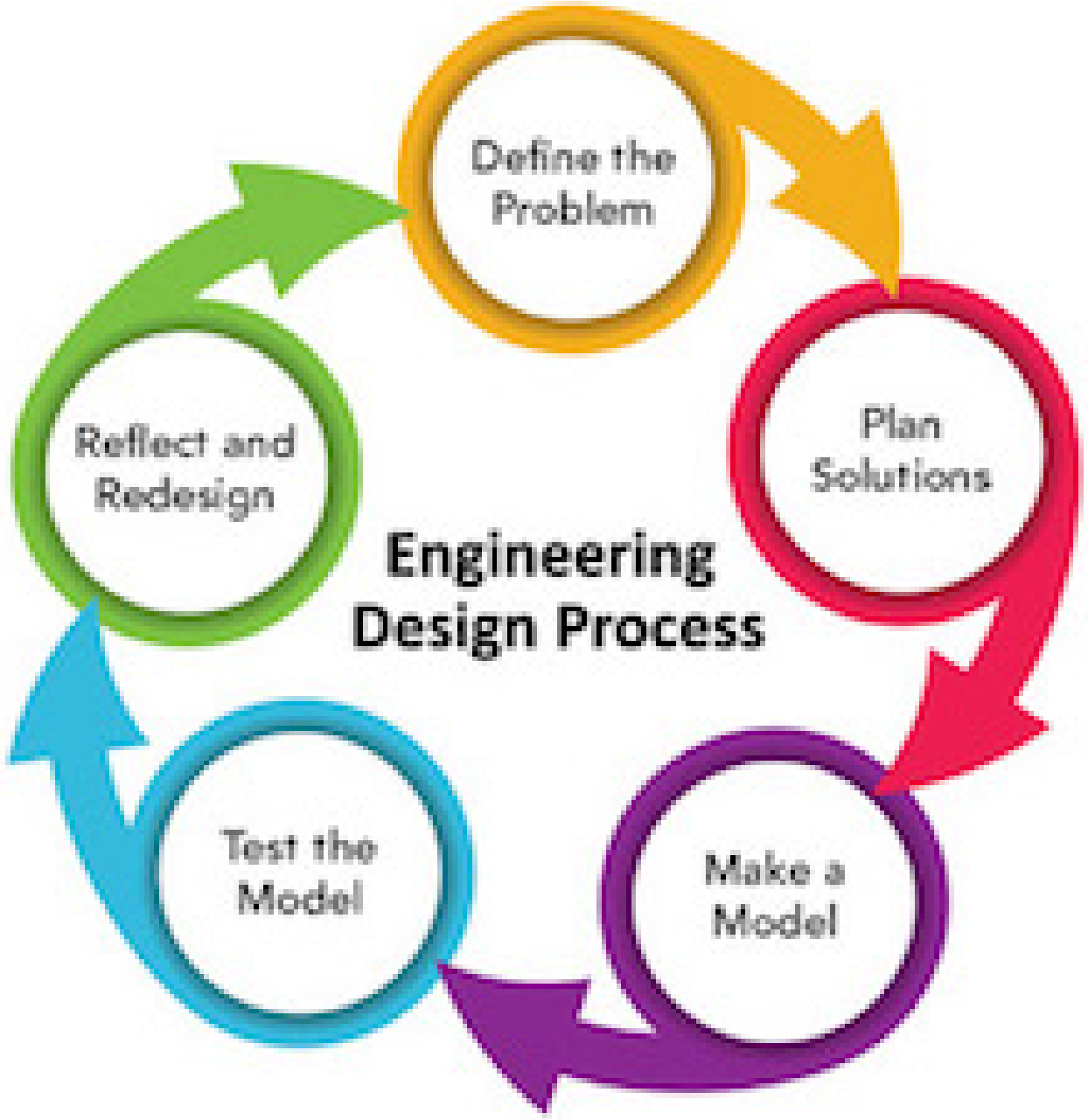
STEM-Practices

- US National Science Foundation
- A curriculum based on the idea of educating students in four specific disciplines — science, technology, engineering and mathematics — in an interdisciplinary and applied approach. Live Science



STEM image

Big Idea of STEM: Design, Model, Test, Re-Design



Middleweb.com
Stem Chart

What STEM Looks Like in Other Domains

Stem Chart

STEM Practices

SCIENCE	ENGINEERING	TECHNOLOGY	MATHEMATICS
Ask questions	Define problems	Become aware of the web of technological systems on which society depends	Make sense of problems and persevere in solving them
Develop and use models	Develop and use models		Model with mathematics
Plan and carry out investigations	Plan and carry out investigations	Learn how to use new technologies as they become available	Use appropriate tools strategically
Analyze and interpret data	Analyze and interpret data		Attend to precision
Use mathematics and computational thinking	Use mathematics and computational thinking	Recognize the role that technology plays in the advancement of science and engineering	Reason abstractly and quantitatively
Construct explanations	Design solutions		Look for and make use of structure
Engage in argument from evidence	Engage in argument from evidence	Make informed decisions about technology given its relationship to society and the environment	Construct viable arguments and critique the reasoning of others
Obtain, evaluate, and communicate information	Obtain, evaluate, and communicate information		Look for and express regularity in repeated reasoning
Sources of the STEM Practices			
Practices of Science & Engineering in Next Generation Science Standards	Practices of Science & Engineering in Next Generation Science Standards	*STEM Lesson Essentials Grade 3-8* Derived from the *Framework for K-12 Science Education*	Mathematical Practices from the Common Core State Standards in Mathematics

GAISE-Recommendations and Goals for Universal Statistics Education

6 Recommendations

1. Teach statistical thinking.
   A. 4 Step Statistical Process
      i. Formulate a question that can answered by data
      ii. Design and implement a plan to collect appropriate data
      iii. Analyze the collected data by graphical and numerical methods
      iv. Interpret the analysis in the context of the original question
2. Focus on conceptual understanding.
3. Integrate real data with a context and a purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

9 Goals, abbreviated

1. Be critical consumers
2. Know when statistics is useful for the investigative process
3. Represent data and interpret graphs and numerical summaries
4. Understand variability
5. Understand randomness
6. Gain experience with statistical models
7. Use statistical inference
8. Draw conclusions using statistical software
9. Aware of ethical issues regarding statistical practice

Why is it important for today's student to have a more in depth understanding of statistics?

- Data Driven Society
  - Goal of collecting and analyzing data is to make us healthier, safer, more productive and efficient
  - Data backfires when it is a priority over people
    - Education-Increased standardized testing in the classroom disenfranchises students
    - Medical-Physician recording data in real time interferes with patient relations
- Daily Headlines about Research
- Data guides policy and decisions in every industry

Teaching Methods ⟶ Student Centered Learning (SCL)

CHARACTERISTICS OF INSTRUCTIONAL ACTIVITIES/TASKS	ORCHESTRATION OF MATHEMATICAL COMMUNICATION
Focus on the "why" as well as the "how"	Focus on the "why" as well as the "how"
Allow for multiple entry points and solution methods	Encourage students to justify and explain their solution strategies
Challenge students to reason about mathematics by looking for patterns, making conjectures, conducting explorations, examining connections between and among mathematical concepts, and justifying mathematical solutions/results	Encourage students to critique the mathematical reasoning of others
Make explicit the connections between mathematics and real-life experiences	Support students in advancing, but not taking over their thinking as they engage in a productive struggle with mathematics
Encourage the use of different tools, including technology, to explore mathematics and solve mathematics problems	Elicit and make connections between different mathematical ideas and/or approaches to the same problem
Provide opportunities for collaboration to communicate and critique mathematical reasoning	

Nellie Mae Education Foundation

NCTM-Questioning Pattern

Funneling

- Teacher begins with probing questions, followed by higher level Q's
- Students answer superficially because there is not enough time for responses

Focusing

- Teacher strives to push students to clarify ideas and make math visible

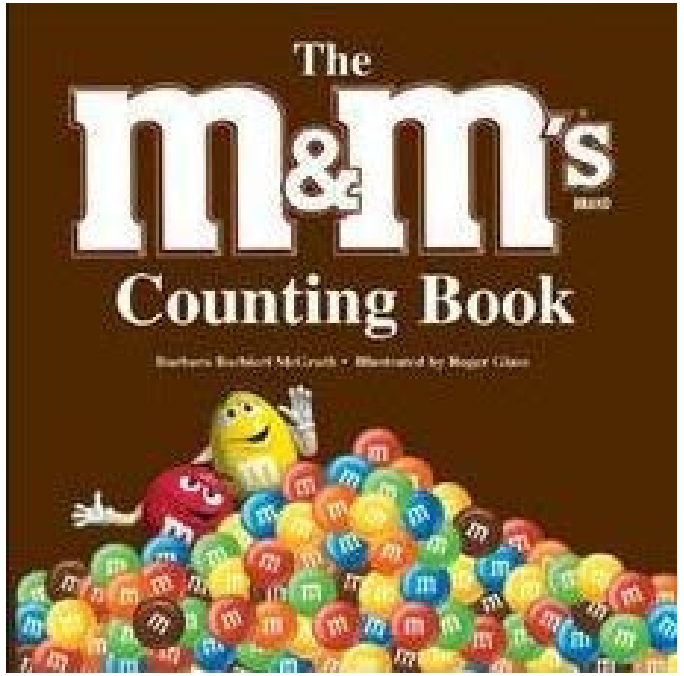
Questioning pattern: Funneling	Questioning pattern: Focusing
T: What do you notice about the graph? [*waits briefly*] Do you see a pattern in the data? [*waits briefly again*] What are the measures of center for the pennies?	*T:* What things do you notice or wonder about the age of pennies?
S1: The mean is about 12.9 years, and the median is about 9 years.	*S1:* It doesn't seem like many of them are very old.
T: What does the box plot tell us about the variability of the data?	*T:* What about the graph makes you say that?
S2: It has a long tail on one side.	*S1:* There's a big mound for newer pennies.
T: That may be true, but what about the interquartile range—the IQR? What does it tell us?	*T:* Is there anything else that you notice?
S3: Where most of the pennies occur.	*S2:* I found the interquartile range and saw that most pennies are from 3 to 19 years old.
T: Is that really what the IQR tells us? What does each part of the box plot stand for?	*T:* Explain to us what the interquartile range tells us.
	S2: It is where most of the pennies occur.
	T: What do you mean by "most of the pennies"?

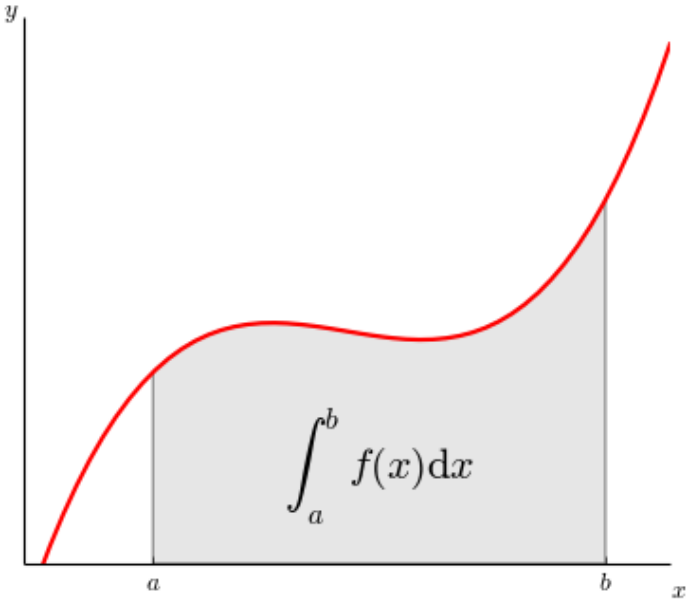
Student Centered

- Students report being more engaged and interested
- Student score higher on PISA (Program International Student Achievement) type questions (NME Foundation)
  - OECD.org (Organization for Economic Cooperation and Development)
- Blend of teaching styles
- Aligns with STEM

Big Picture of Statistics

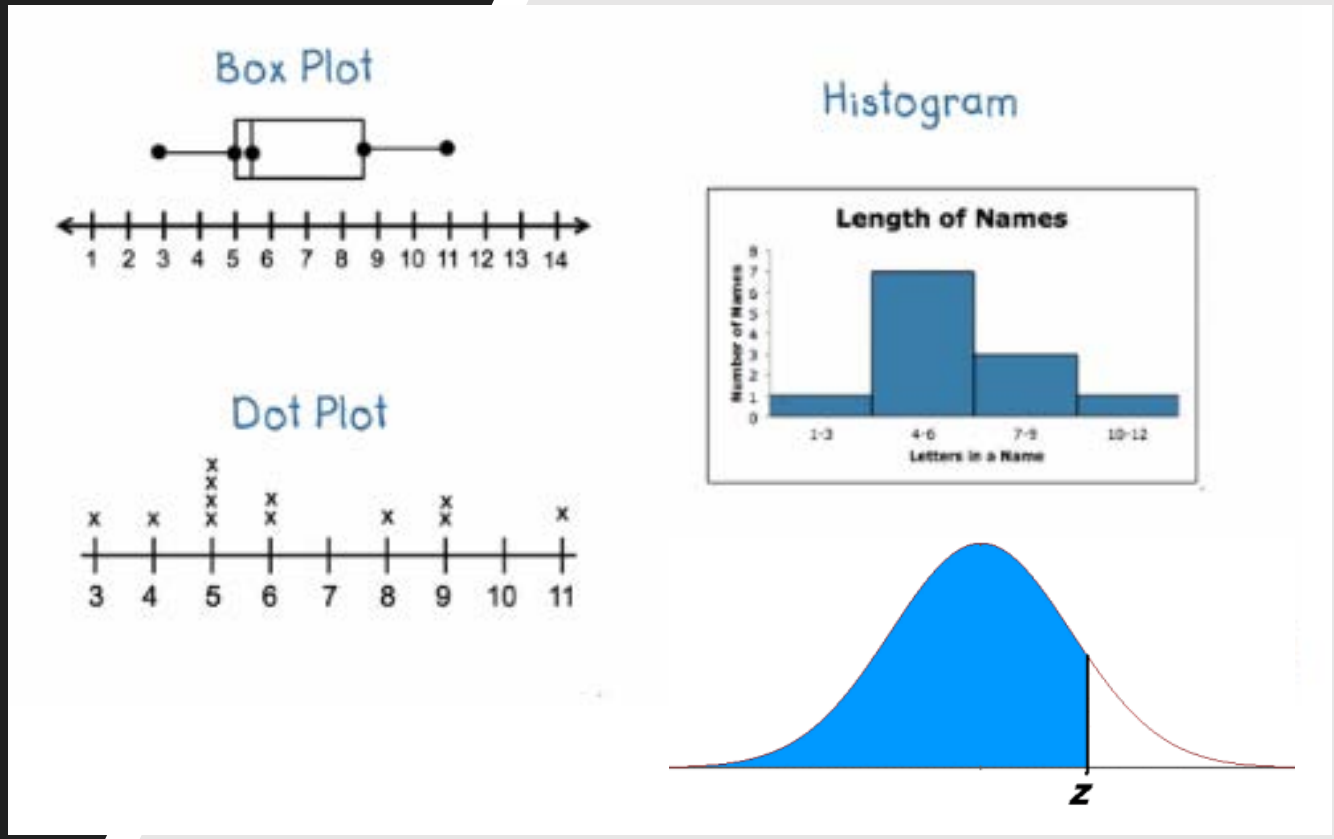
- From Pre-school (Counting M&M's) to Graduate School (Modeling and Integrating a Curve)





Graphical Representation of Data

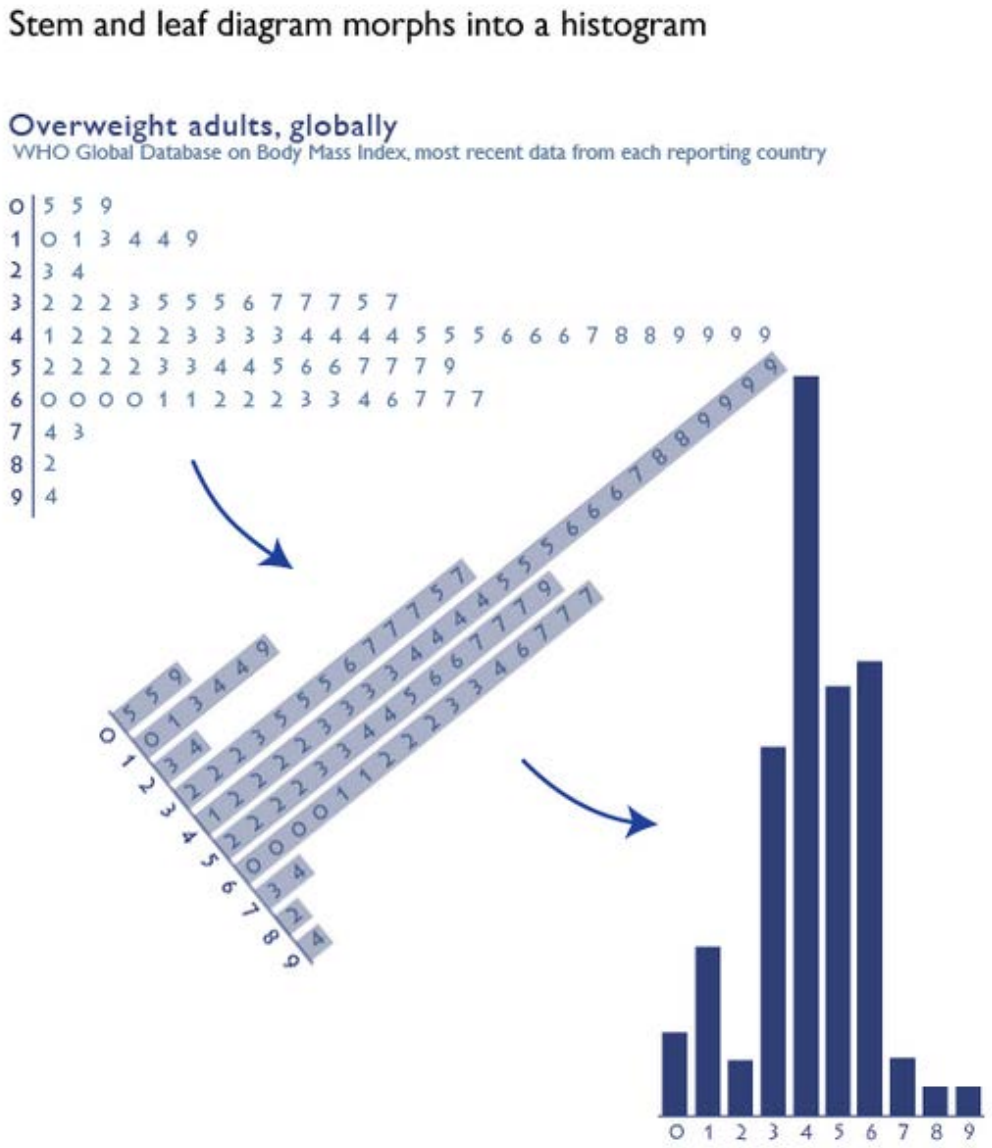
- Table
- Dot Plot
- Stem-Leaf
- Box Plot
- Histogram
- Curves



Khan Academy resource:
[Comparing Data Displays]

Image Source: Learn Zillion

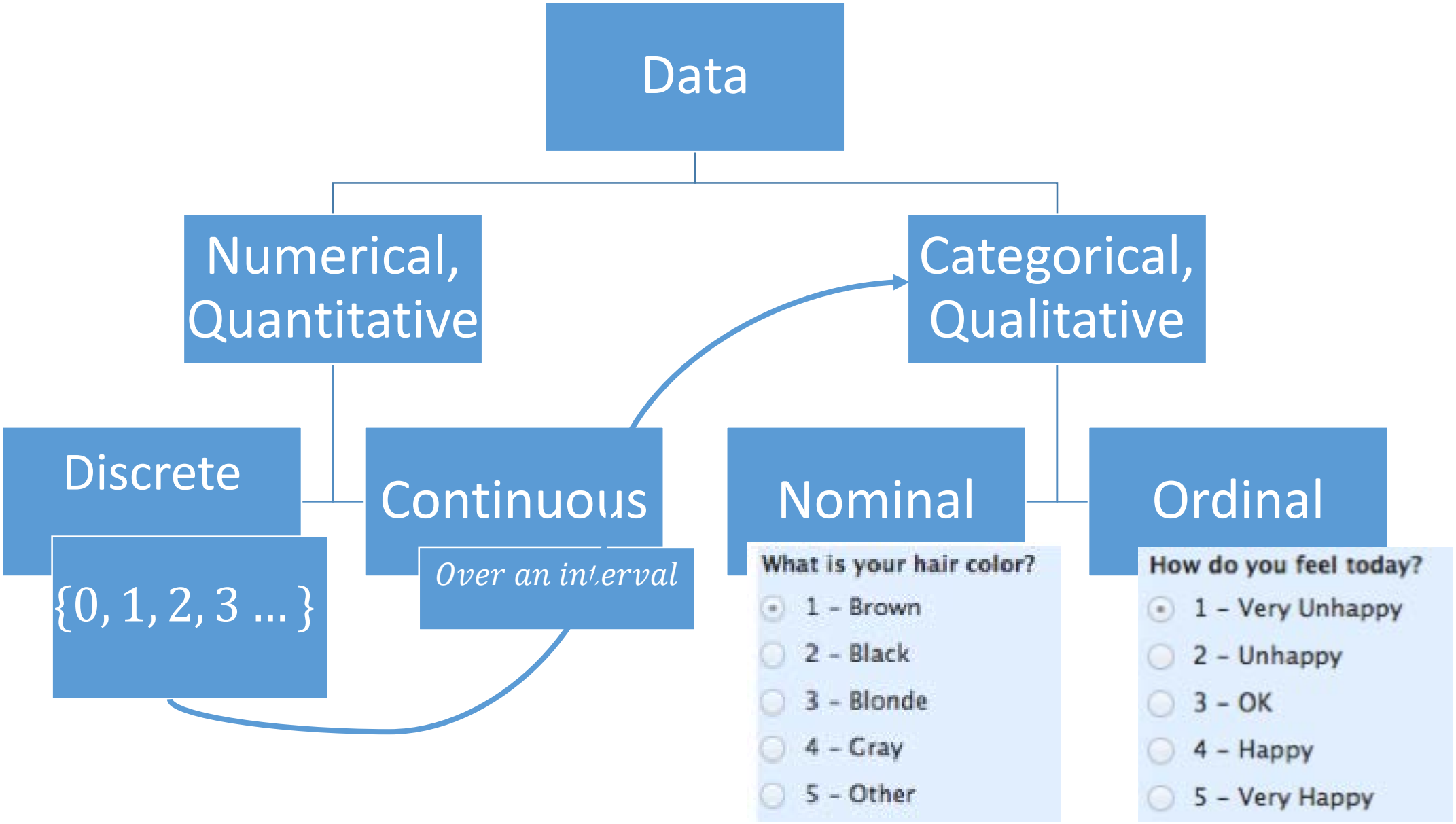
Stem and Leaf morphs into a Histogram



Stem-Leaf

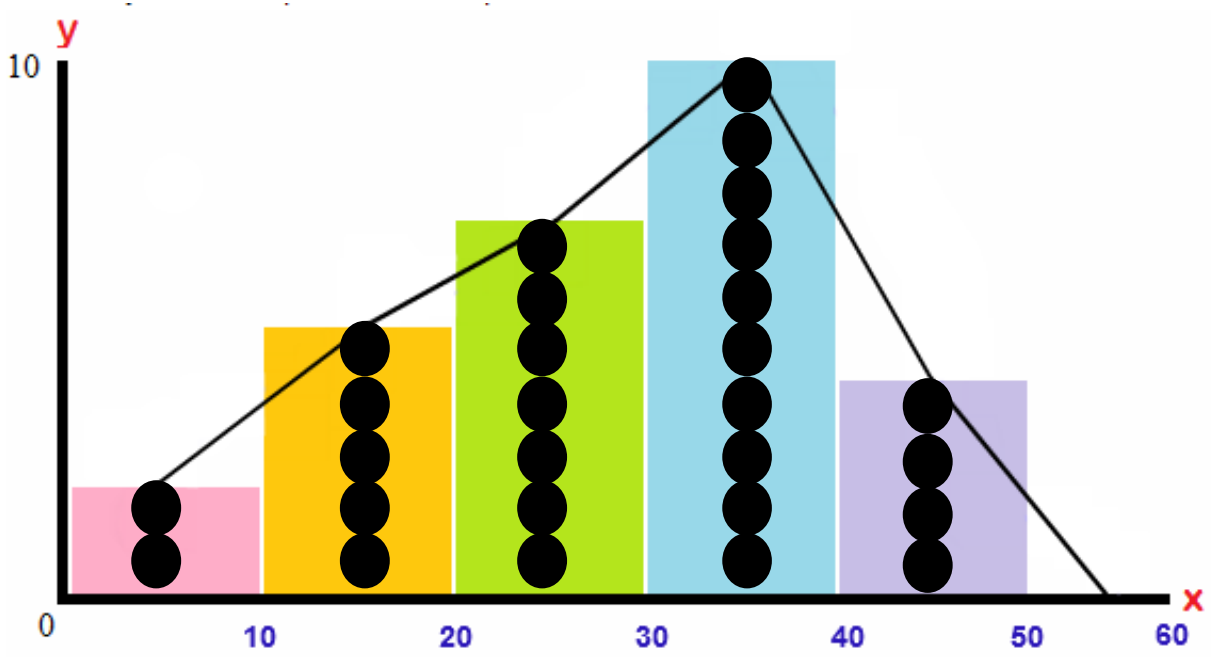
Stem	Leaf
22	8
23	1, 2, 5, 6, 7
24	3, 4, 7, 9, 9
25	0, 3, 5, 8, 8, 9
26	2, 2, 3, 5, 9

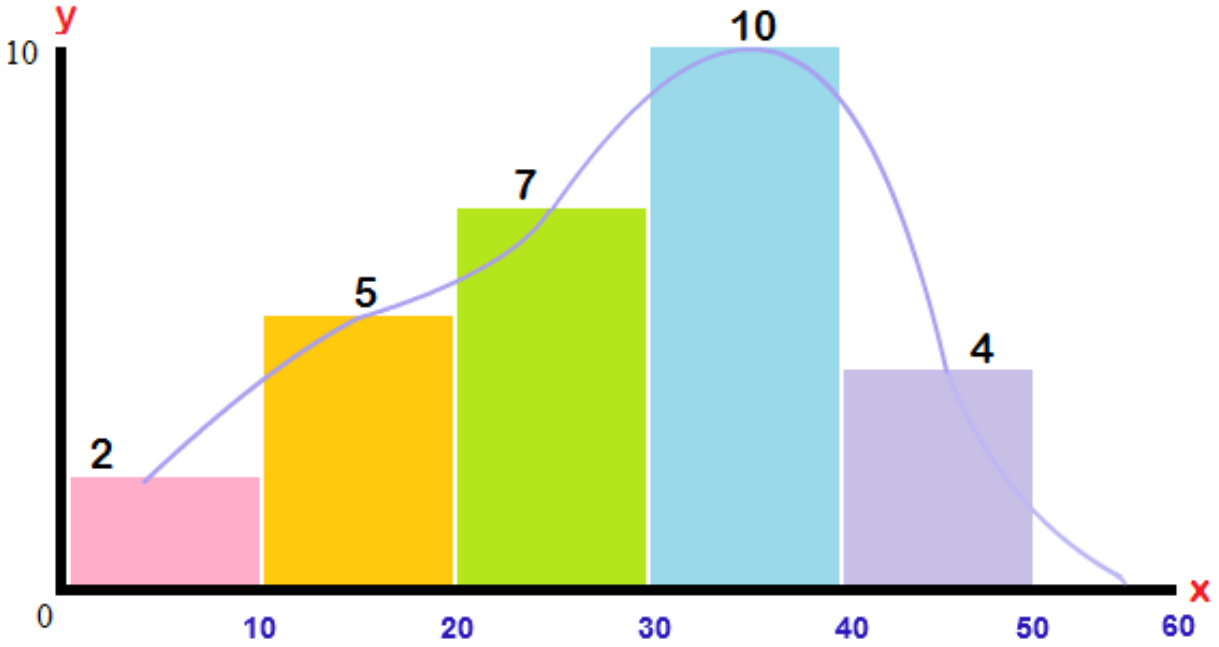
Key 22|8 is SAT score 228


Data
Numerical, Quantitative
Categorical, Qualitative
Discrete
{0, 1, 2, 3 … }
Continuous
Over an interval
Nominal
What is your hair color?
1 – Brown
2 – Black
3 – Blonde
4 – Gray
5 – Other
Ordinal
How do you feel today?
1 – Very Unhappy
2 – Unhappy
3 – OK
4 – Happy
5 – Very Happy

Distribution of Data

Dot Plots → Histograms → Frequency Polygons → Curves





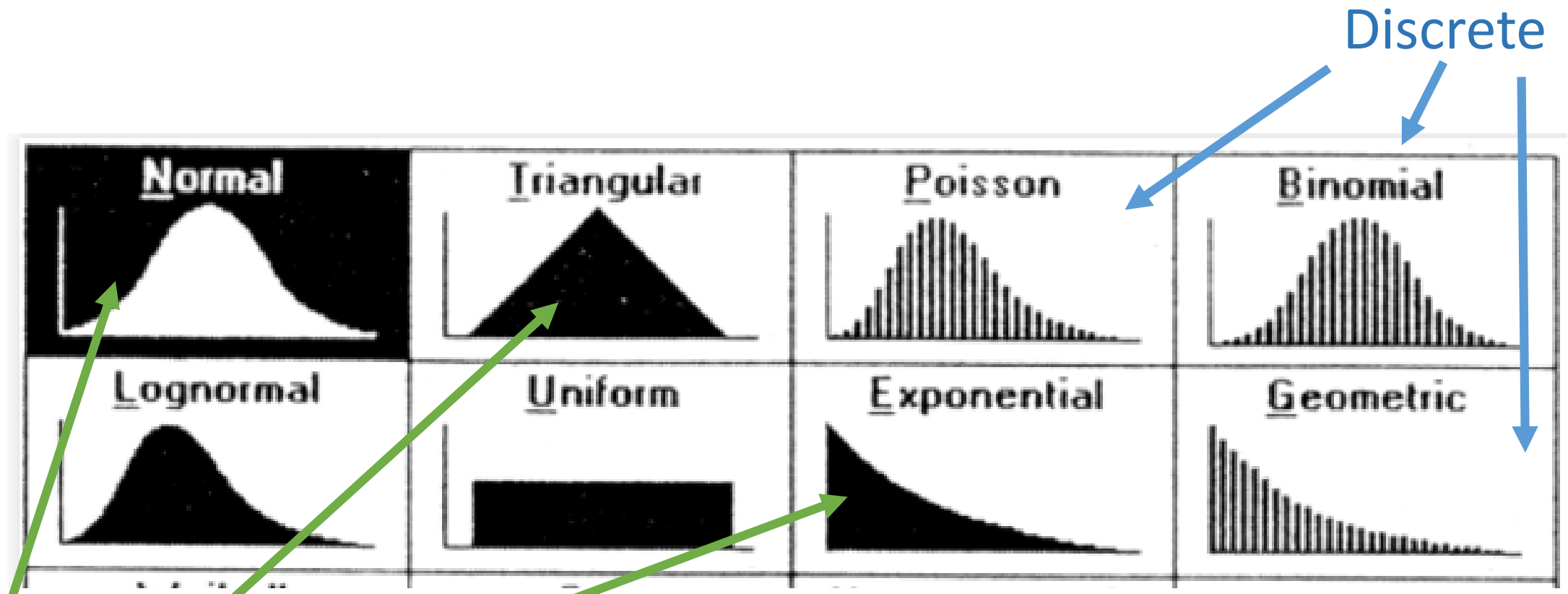
Where is the bulk of the data? (Funnel!)

What to you notice about the graph? (Focus!)

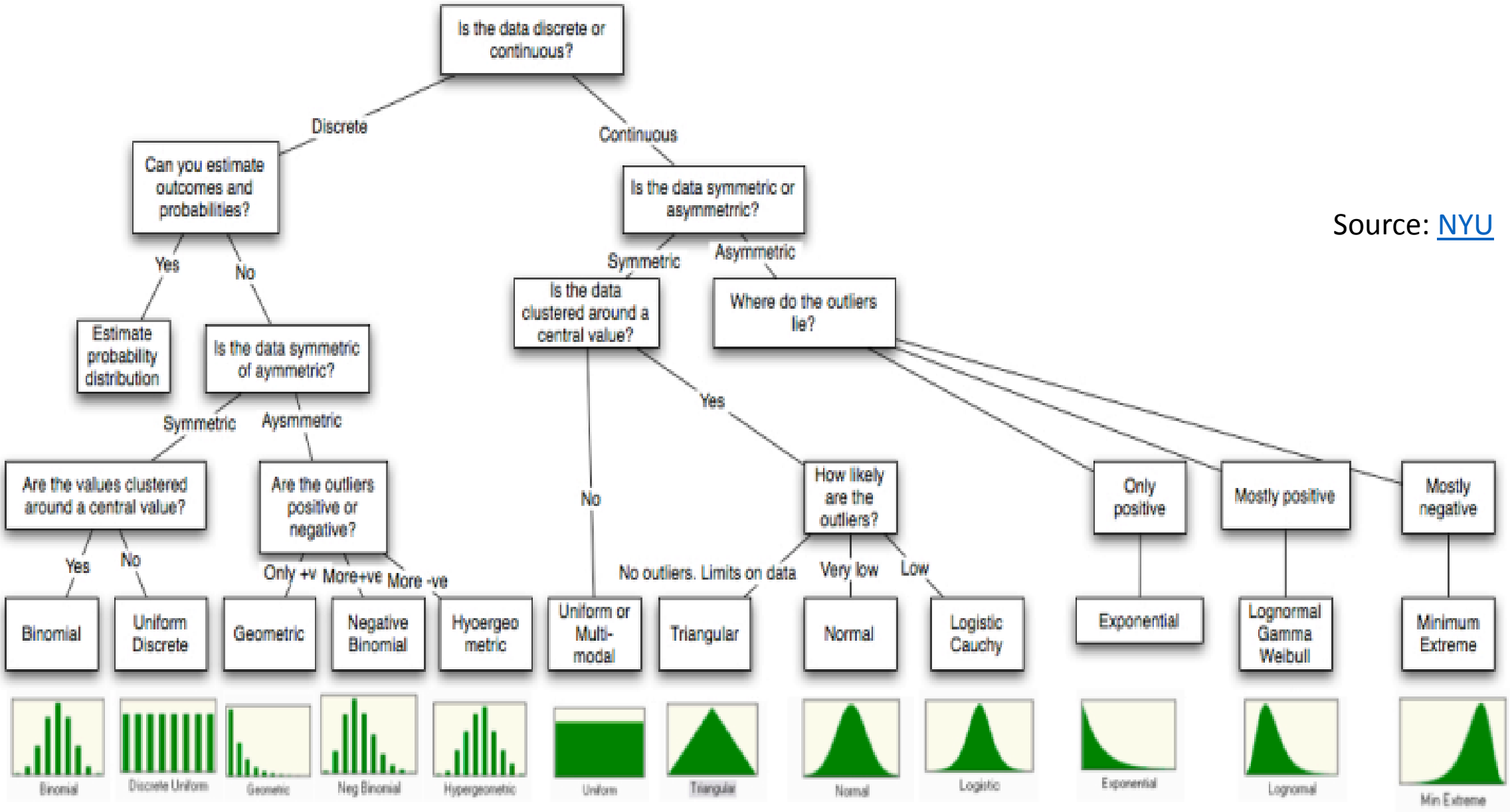
What percent of the data is less than 30?

What is the probability that a randomly selected unit is less than 30, $P(X < 30)$?

Common Distributions, Probability Density Functions (pdf's)



Modeling Data with Distributions*



Source: NYU

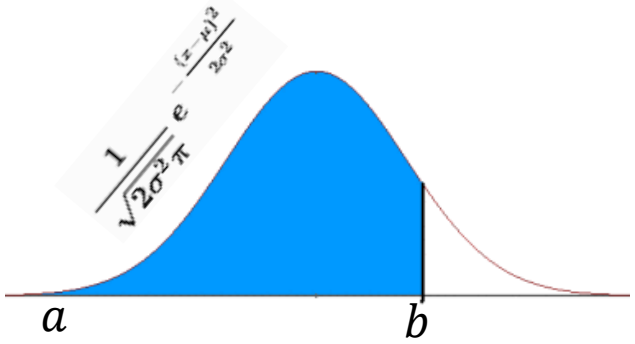
Probability Distributions

- Each Distribution (histogram or curve) is defined by a function, a model
- Discrete Data, take the sum of the model: $\sum_{i=1}^{n} p_i \cdot x_i$
- Continuous, integrate the model: $\int_a^b f(x)\,dx$

- Example Distribution: $N\sim(\mu, \sigma)$
  - Parameters mean, μ, and standard deviation, σ.
- Equation for the normal curve: $f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Its continuous, so integrate to find the area under the curve,

$$\int_a^b \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

- The area under the curve represents the probability for that event



Relative Frequency

- $y - axis$ represents percent or proportion instead of actual count
- Same shape as Frequency Histograms

Cumulative

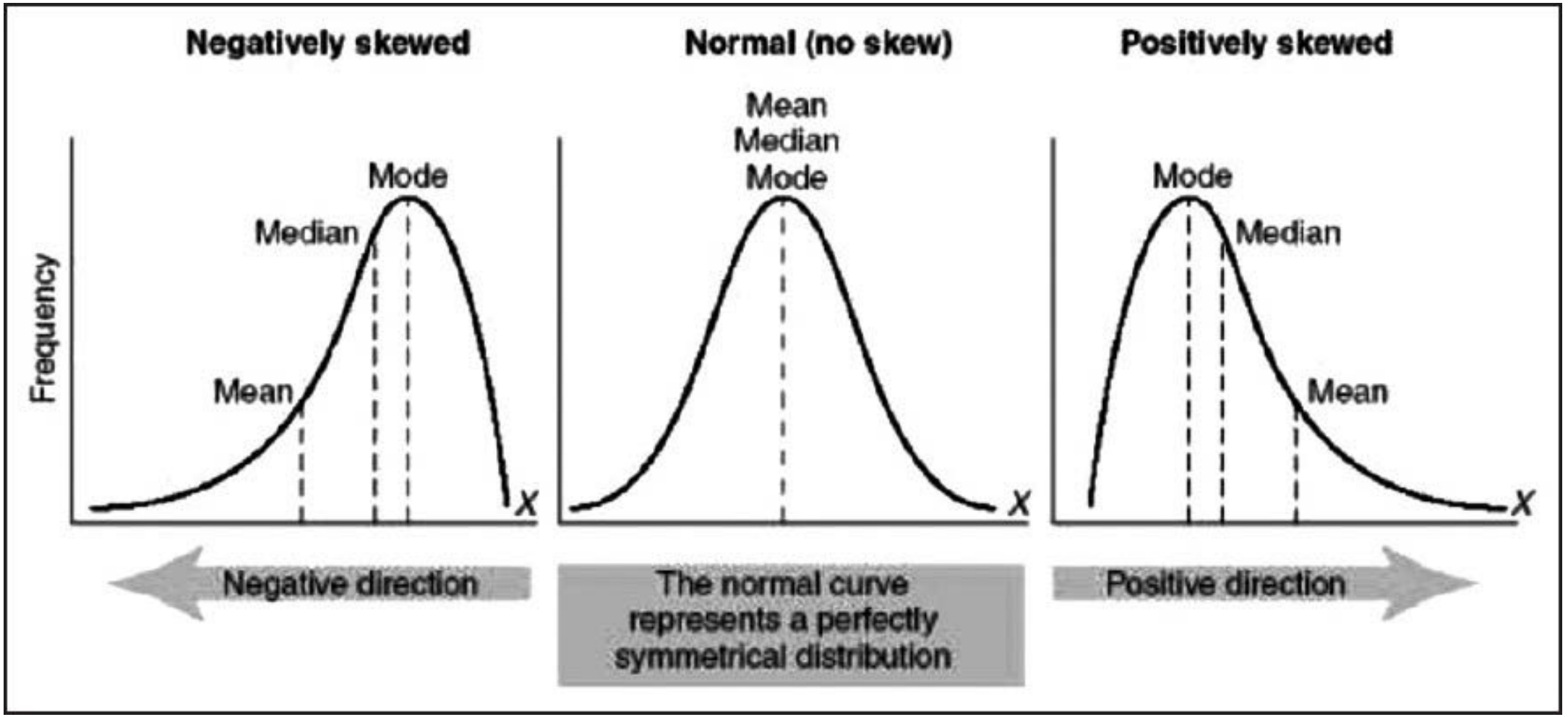
Pareto

Describe how the data is distributed (SOCS)

- **S**hape
- **O**utliers
  - An observation that outside the pattern or distribution
- **C**enter
- **S**pread

Source: Mathworld Wolfram

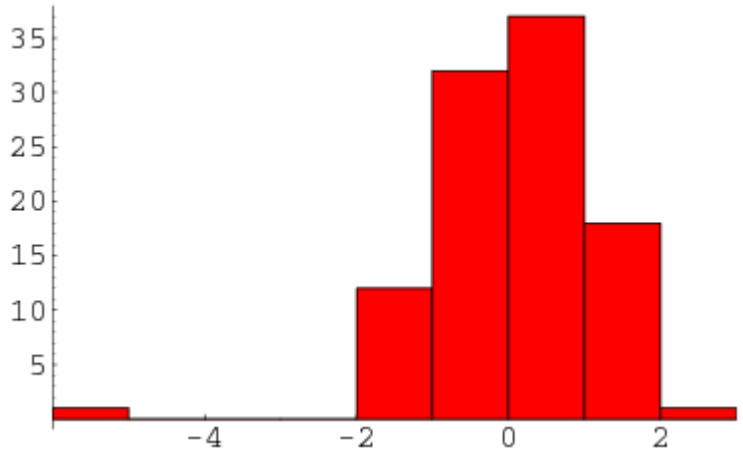
Shape



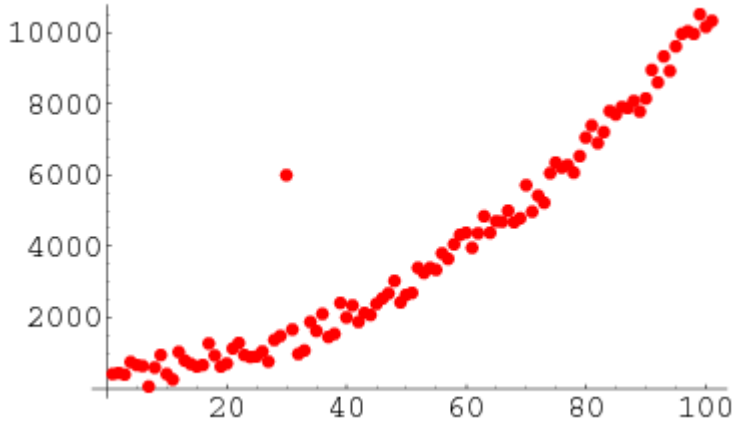
Source: **Descriptive statistics. Available from** http://www.southalabama.edu

Identifying Outliers

Histogram



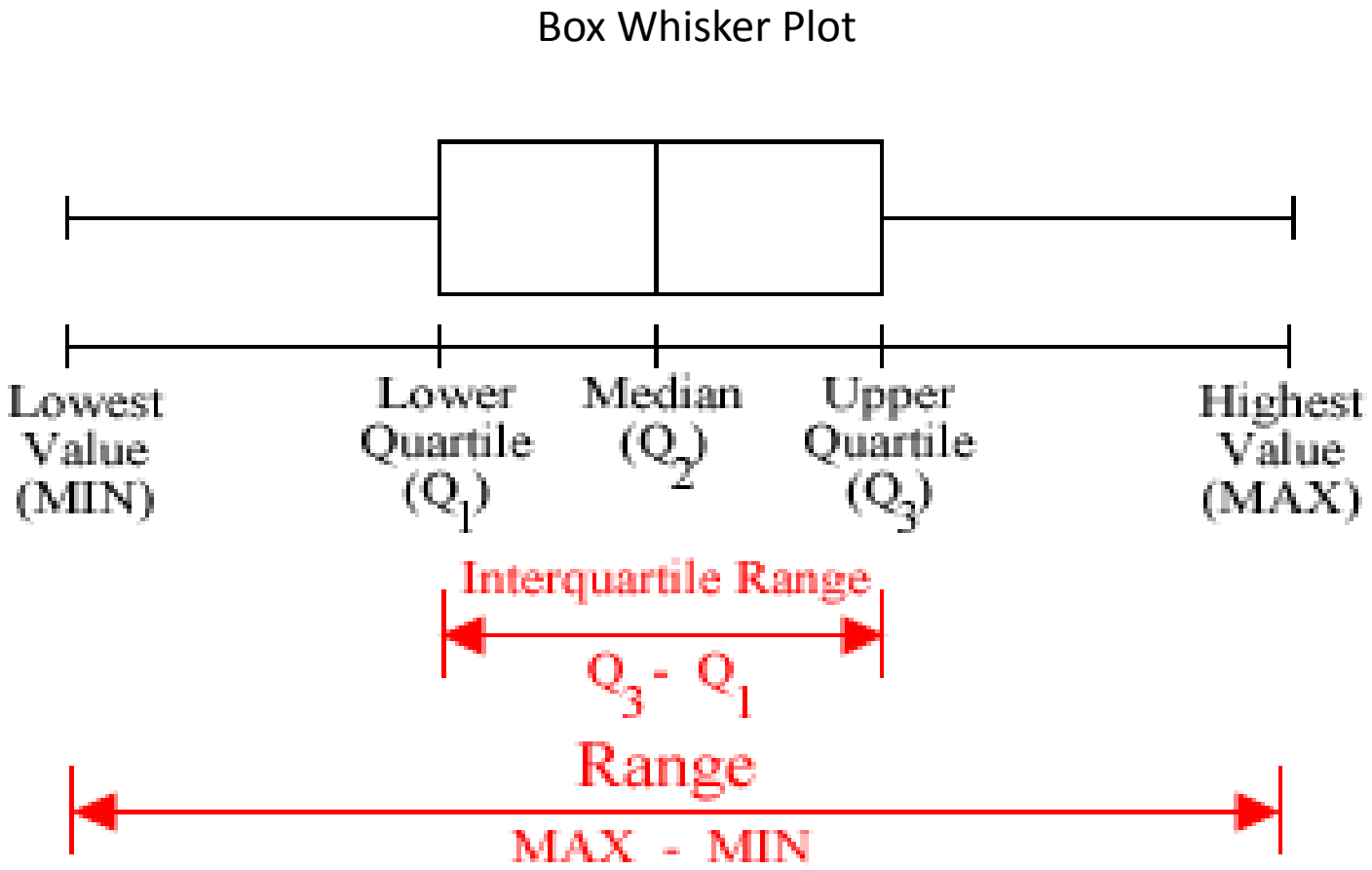
Linear Regression



Source: [Mathworld Wolfram]

Computing Outliers

- $Q1 - 1.5(IQR)$
- $Q3 + 1.5(IQR)$



Descriptive Statistics

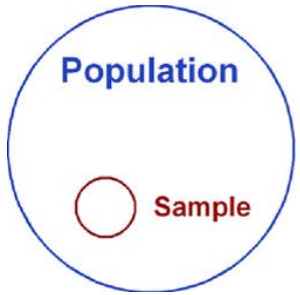
Center

- Mean
- Median
- Mode

Spread – Variability, Dispersion

- Standard Deviation
- Variance
- Range
- Interquartile Range

Statistics Notation



Data	Population	Sample
• Mean	• μ	• $\bar{x}$
• Standard Deviation	• σ	• s
• Number of data points	• N	• n
• Characteristic being measured	• Parameter	• Statistic

Formula for Mean

Same Formula

Sample Mean	Population Mean
$\bar{x} = \frac{\Sigma x}{n}$	$\mu = \frac{\Sigma x}{N}$

where Σx is sum of all data values

N is number of data items in population

n is number of data items in sample

Formulas for Standard Deviation

Different Formula

Sample

- $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}}$

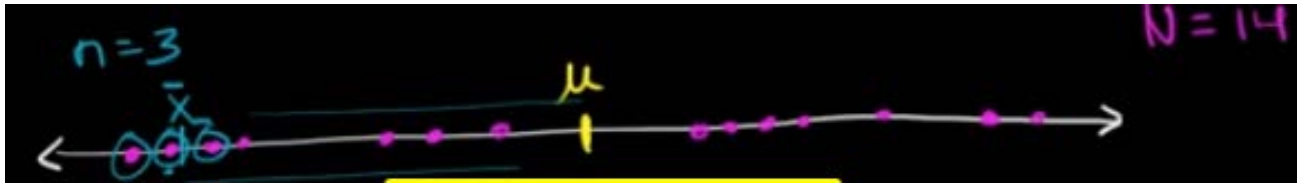
Population

- $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)}{N}}$

- Variance
  - $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)}{N}$

Why is the Standard Deviation of the Sample divided by (n-1)?*

- Dividing (n-1) instead of n
  - Gives an unbiased, more accurate estimate, of the population.
  - A larger value
- If you divide by n,
  - Then you will get a smaller value and likely to underestimate the standard deviation, distance from the mean.



Note: Population Mean and Sample Mean are the same formula because the Sample Mean will always be within your sample.

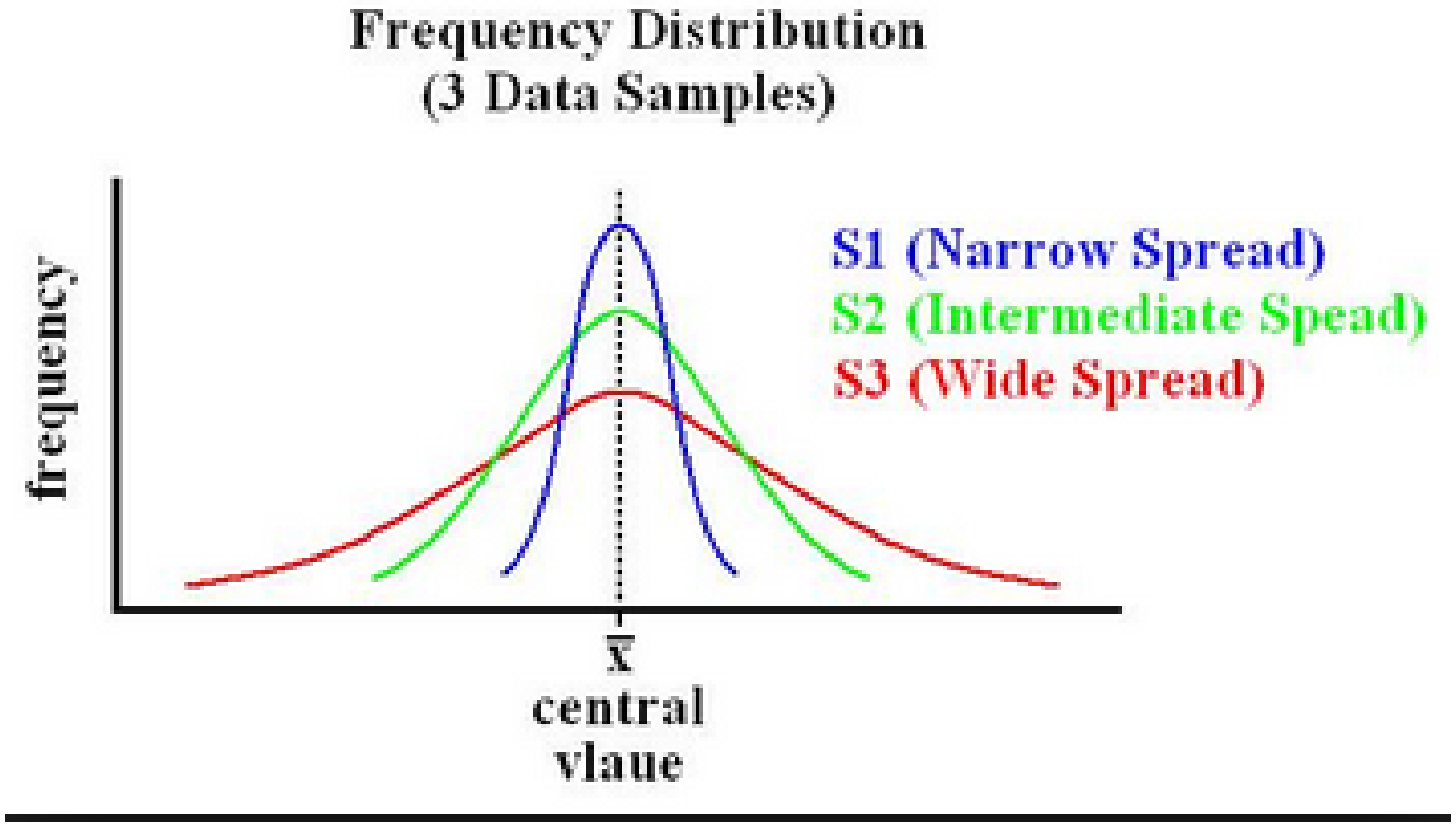
Khan Academy, why (n-1)?

Khan Academy, Simulation for more intuition on (n-1).

Intuition behind Standard Deviation, Variability
(Variation or Spread)

- Make a Dot Plot of Shoe Size on the board
- Unimodal or Bimodal?
- Narrow data to women's shoe size
- Compute the sample mean, $\bar{x}$, and label on dot plot
- Draw the distances from each data point to the mean
- Square each distance
- Compute the average of the distances
- Take the square root

SPREAD-Variability



Statistics for Middle and High School Teachers

A Resource for Middle and High School Teachers to Feel Better Prepared to Teach the Common Core State Standards (CCSS) relating to Statistics.

Welcome!

- Pick up a binder and Tinspire calculator
- Fill out name tags
- Find our online classroom at Schoology.com
  - Class code:
  - Begin Survey Monkey
- Introductions
  - Name
  - Education experience
  - A hobby or interest

Workshop Goal: Prepare Middle and High School Teachers to Teach Statistics

- Review and Learn Statistics through Student-Centered Problems
- Represent and analyze data with the software and technology
  - TInspire
  - Excel
  - Google Sheets
  - Khanacademy.com
  - Wolfram Mathematica
  - Desmos.com
  - Geogebra.com

Workshop Outline

- Mathematics Curriculum and Practices Today
- Representing Data
- Experimental Design
- Normal Distribution
- Linear Regression

Mathematics Curriculum and Practices Today

- (CCSS) Common Core State Standards
- (NCTM) National Council of Teaches of Mathematics
- (STEM) Science Technology Engineering and Mathematics
  - Concept began in meetings sponsored by the National Science Foundation (NSA)
- (GAISE) Guidelines for Assessment and Instruction in Statistics Education
  - By American Statisticians Association (ASA)
- Connection
  - The CCSS derived from the NCTM standards which were developed in 1989 and updated in 2000. The NCTM standards were praised in the education and science industries. The National Science Foundations funded projects based on these standards. STEM came out of meetings held by the National Science Foundation and started showing up in the early 2000's.

CCSSM-Probability and Statistics

Standards

- Develop understanding of statistical variability.
- Summarize and describe distributions.
- Use random sampling to draw inferences about a population.
- Draw informal comparative inferences about two populations.
- Investigate chance processes and develop, use, and evaluate probability models.
- Investigate patterns of association in bivariate data.
- Interpreting Categorical and Quantitative Data*
- Making Inferences and Justifying Conclusions*
- Conditional Probability and the Rules of Probability*
- Using Probability to Make Decisions*

8 Mathematical Practices

- Make sense of problems and persevere in solving them.
- Reason abstractly and quantitatively.
- Construct viable arguments and critique the reasoning of others.
- Model with mathematics.
- Use appropriate tools strategically.
- Attend to precision.
- Look for and make use of structure.
- Look for and express regularity in repeated reasoning.

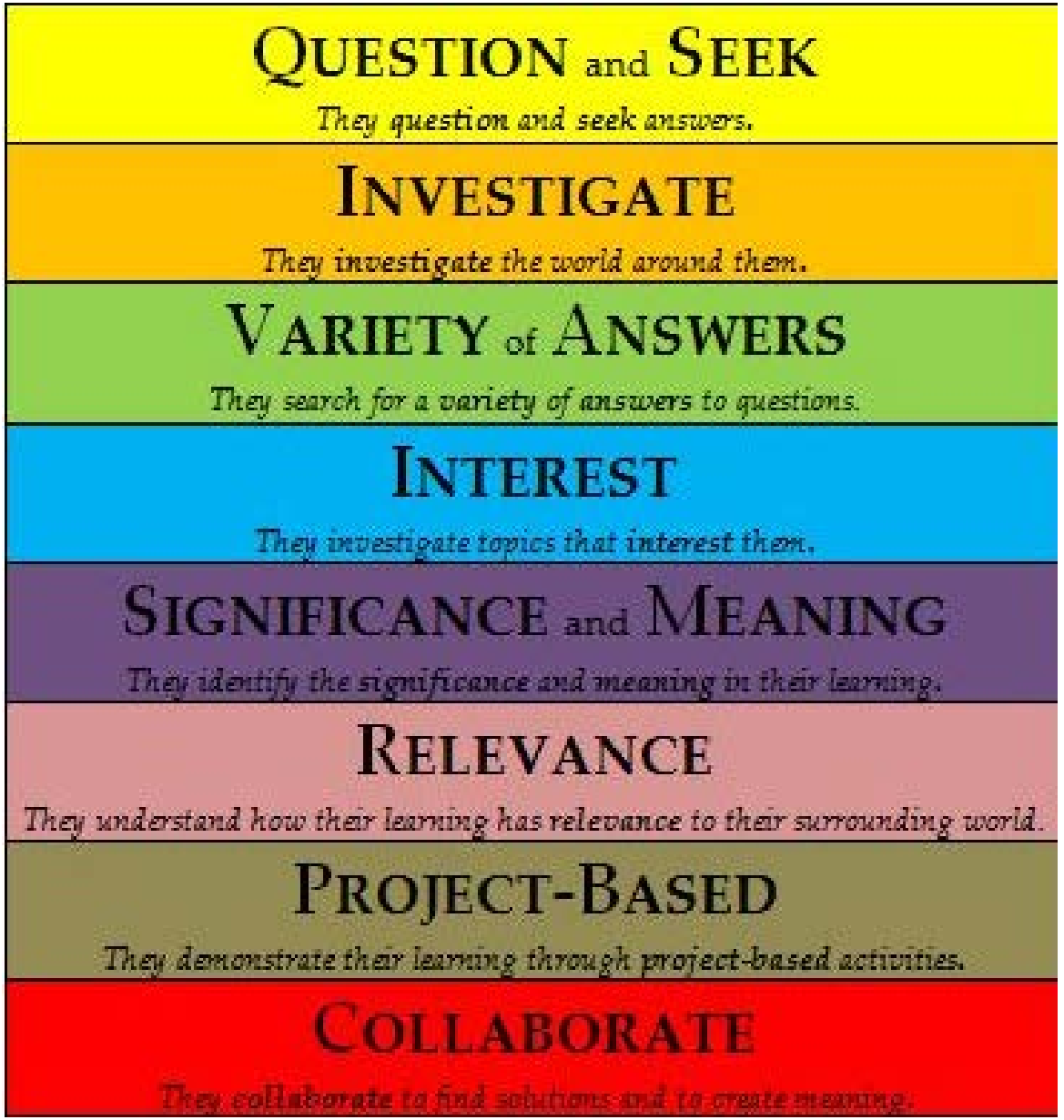
NCTM – Data Analysis and Probability

- **Summary of Standards and Principles**
  - Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them
  - Select and use appropriate statistical methods to analyze data
  - Develop and evaluate inferences and predictions that are based on data
  - Understand and apply basic concepts of probability

NCTM-Detailed Data Analysis and Probability Standards

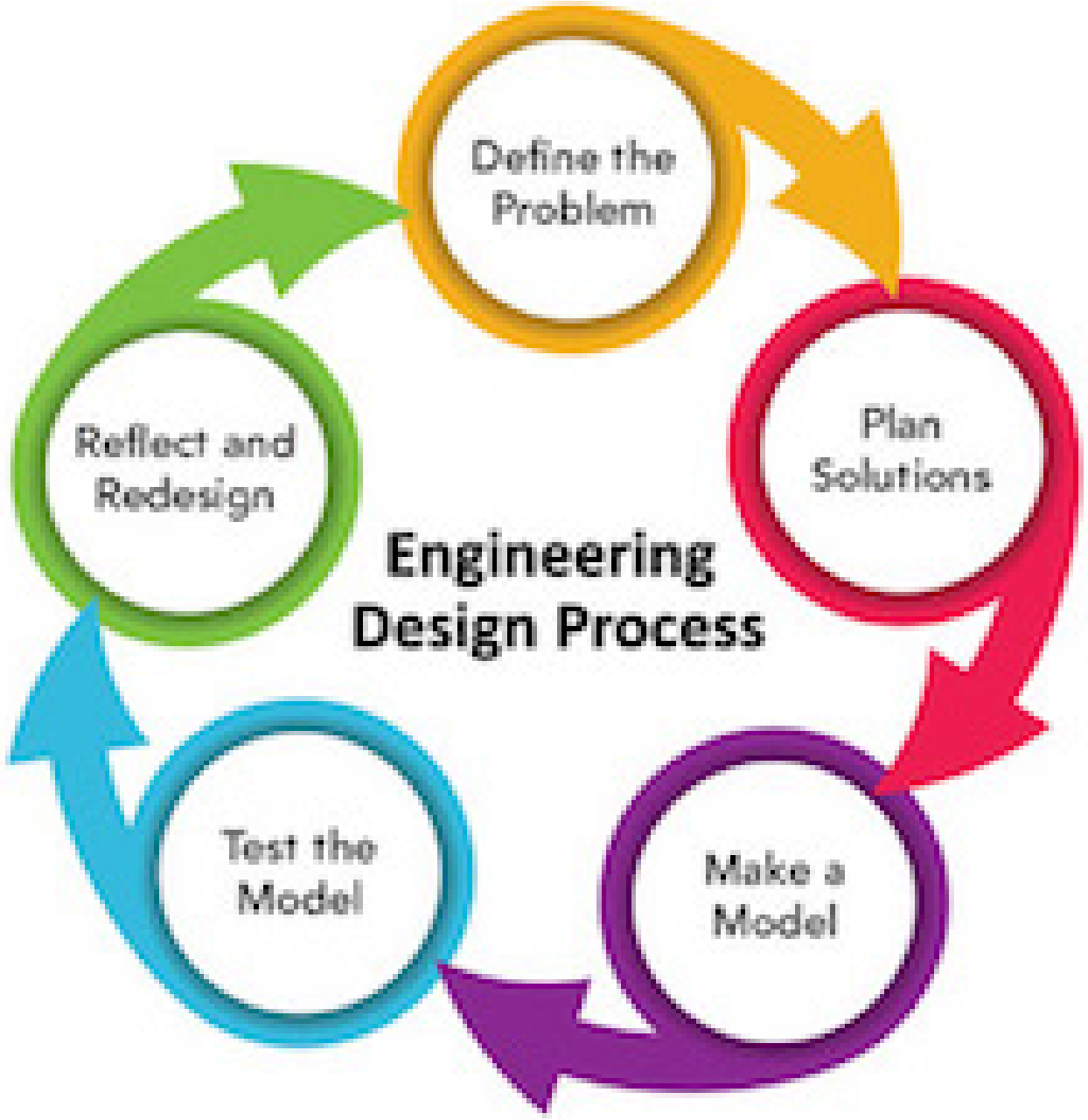
STEM-Practices

- US National Science Foundation
- A curriculum based on the idea of educating students in four specific disciplines — science, technology, engineering and mathematics — in an interdisciplinary and applied approach. Live Science



STEM image

Big Idea of STEM: Design, Model, Test, Re-Design



Middleweb.com
Stem Chart

What STEM Looks Like in Other Domains

Stem Chart

STEM Practices

SCIENCE	ENGINEERING	TECHNOLOGY	MATHEMATICS
Ask questions	Define problems	Become aware of the web of technological systems on which society depends	Make sense of problems and persevere in solving them
Develop and use models	Develop and use models		Model with mathematics
Plan and carry out investigations	Plan and carry out investigations	Learn how to use new technologies as they become available	Use appropriate tools strategically
Analyze and interpret data	Analyze and interpret data		Attend to precision
Use mathematics and computational thinking	Use mathematics and computational thinking	Recognize the role that technology plays in the advancement of science and engineering	Reason abstractly and quantitatively
Construct explanations	Design solutions		Look for and make use of structure
Engage in argument from evidence	Engage in argument from evidence	Make informed decisions about technology given its relationship to society and the environment	Construct viable arguments and critique the reasoning of others
Obtain, evaluate, and communicate information	Obtain, evaluate, and communicate information		Look for and express regularity in repeated reasoning
Sources of the STEM Practices			
Practices of Science & Engineering in Next Generation Science Standards	Practices of Science & Engineering in Next Generation Science Standards	*STEM Lesson Essentials Grade 3-8* Derived from the *Framework for K-12 Science Education*	Mathematical Practices from the Common Core State Standards in Mathematics

GAISE-Recommendations and Goals for Universal Statistics Education

6 Recommendations

1. Teach statistical thinking.
   A. 4 Step Statistical Process
      i. Formulate a question that can answered by data
      ii. Design and implement a plan to collect appropriate data
      iii. Analyze the collected data by graphical and numerical methods
      iv. Interpret the analysis in the context of the original question
2. Focus on conceptual understanding.
3. Integrate real data with a context and a purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

9 Goals, abbreviated

1. Be critical consumers
2. Know when statistics is useful for the investigative process
3. Represent data and interpret graphs and numerical summaries
4. Understand variability
5. Understand randomness
6. Gain experience with statistical models
7. Use statistical inference
8. Draw conclusions using statistical software
9. Aware of ethical issues regarding statistical practice

Why is it important for today's student to have a more in depth understanding of statistics?

- Data Driven Society
  - Goal of collecting and analyzing data is to make us healthier, safer, more productive and efficient
  - Data backfires when it is a priority over people
    - Education-Increased standardized testing in the classroom disenfranchises students
    - Medical-Physician recording data in real time interferes with patient relations
- Daily Headlines about Research
- Data guides policy and decisions in every industry

Teaching Methods ⟶ Student Centered Learning (SCL)

CHARACTERISTICS OF INSTRUCTIONAL ACTIVITIES/TASKS	ORCHESTRATION OF MATHEMATICAL COMMUNICATION
Focus on the "why" as well as the "how"	Focus on the "why" as well as the "how"
Allow for multiple entry points and solution methods	Encourage students to justify and explain their solution strategies
Challenge students to reason about mathematics by looking for patterns, making conjectures, conducting explorations, examining connections between and among mathematical concepts, and justifying mathematical solutions/results	Encourage students to critique the mathematical reasoning of others
Make explicit the connections between mathematics and real-life experiences	Support students in advancing, but not taking over their thinking as they engage in a productive struggle with mathematics
Encourage the use of different tools, including technology, to explore mathematics and solve mathematics problems	Elicit and make connections between different mathematical ideas and/or approaches to the same problem
Provide opportunities for collaboration to communicate and critique mathematical reasoning	

Nellie Mae Education Foundation

NCTM-Questioning Pattern

Funneling

- Teacher begins with probing questions, followed by higher level Q's
- Students answer superficially because there is not enough time for responses

Focusing

- Teacher strives to push students to clarify ideas and make math visible

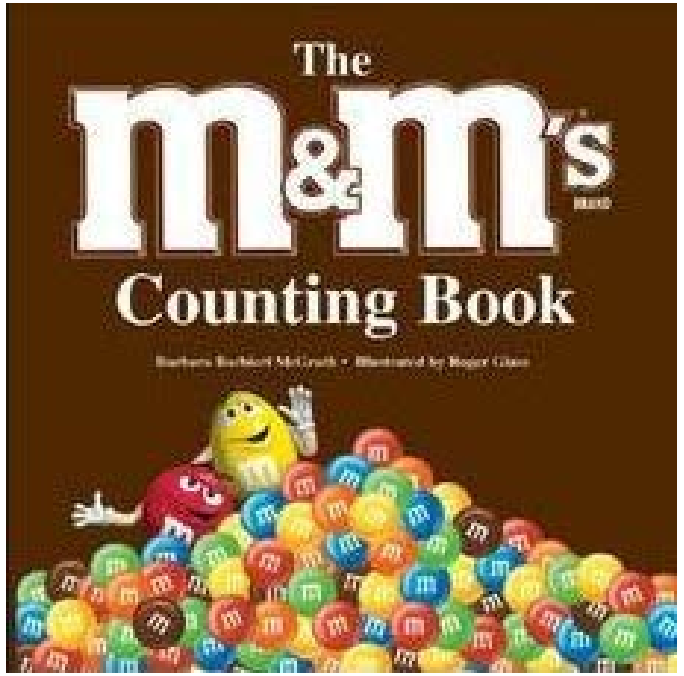
Questioning pattern: Funneling	Questioning pattern: Focusing
T: What do you notice about the graph? [*waits briefly*] Do you see a pattern in the data? [*waits briefly again*] What are the measures of center for the pennies?	*T:* What things do you notice or wonder about the age of pennies?
S1: The mean is about 12.9 years, and the median is about 9 years.	*S1:* It doesn't seem like many of them are very old.
T: What does the box plot tell us about the variability of the data?	*T:* What about the graph makes you say that?
S2: It has a long tail on one side.	*S1:* There's a big mound for newer pennies.
T: That may be true, but what about the interquartile range—the IQR? What does it tell us?	*T:* Is there anything else that you notice?
S3: Where most of the pennies occur.	*S2:* I found the interquartile range and saw that most pennies are from 3 to 19 years old.
T: Is that really what the IQR tells us? What does each part of the box plot stand for?	*T:* Explain to us what the interquartile range tells us.
	S2: It is where most of the pennies occur.
	T: What do you mean by "most of the pennies"?

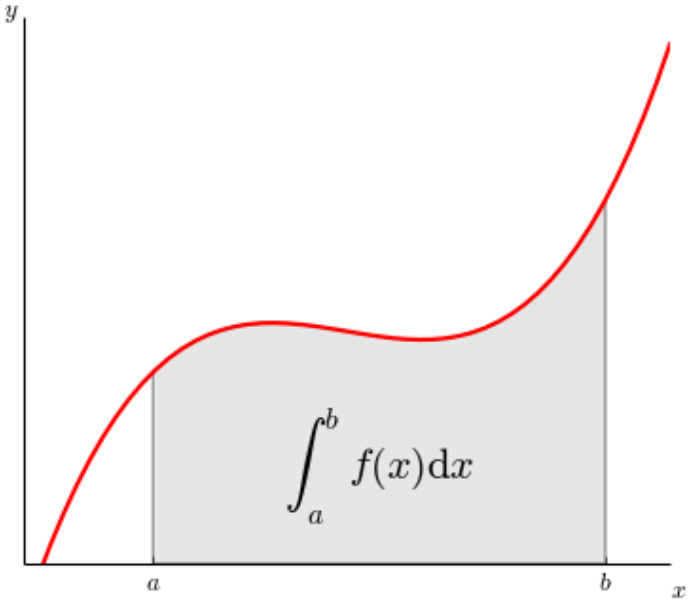
Student Centered

- Students report being more engaged and interested
- Student score higher on PISA (Program International Student Achievement) type questions (NME Foundation)
  - OECD.org (Organization for Economic Cooperation and Development)
- Blend of teaching styles
- Aligns with STEM

Big Picture of Statistics

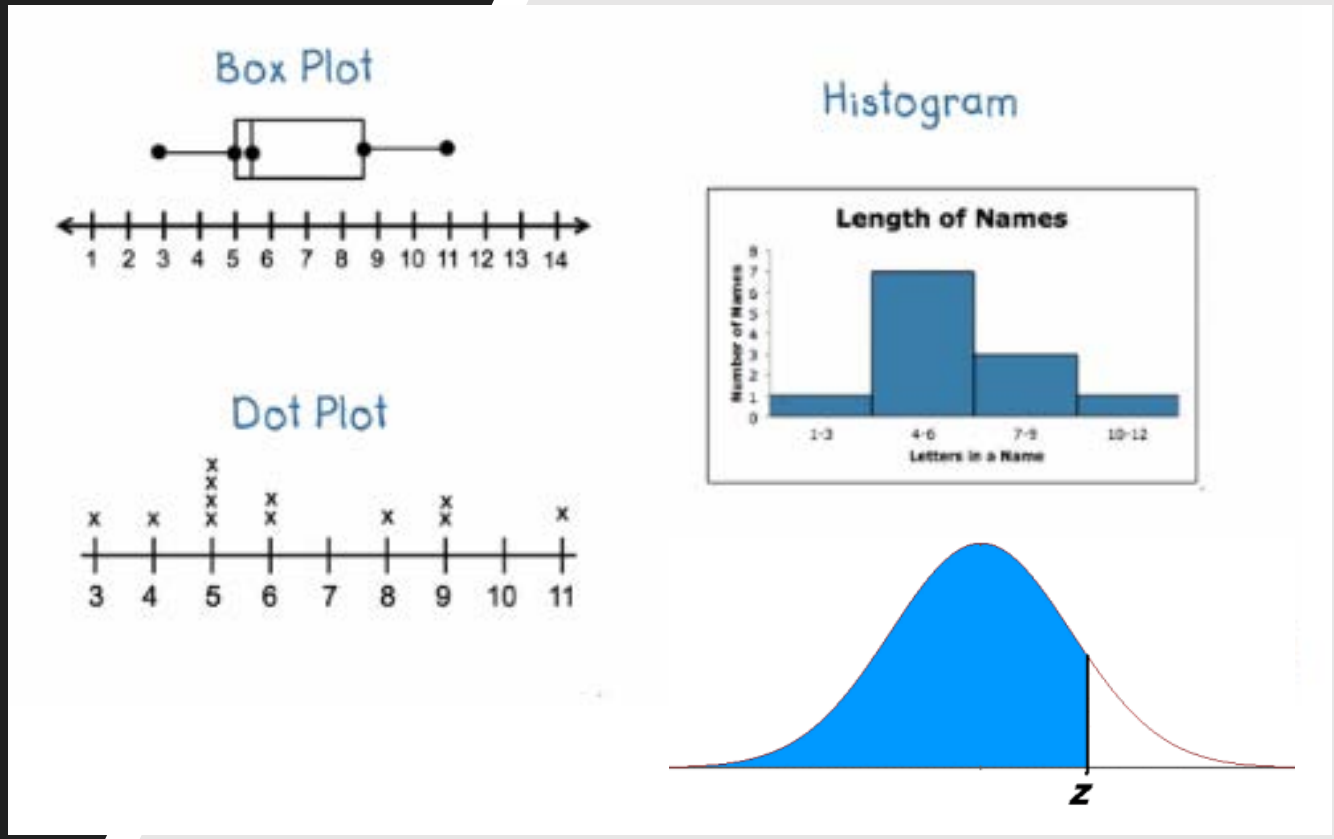
- From Pre-school (Counting M&M's) to Graduate School (Modeling and Integrating a Curve)





Graphical Representation of Data

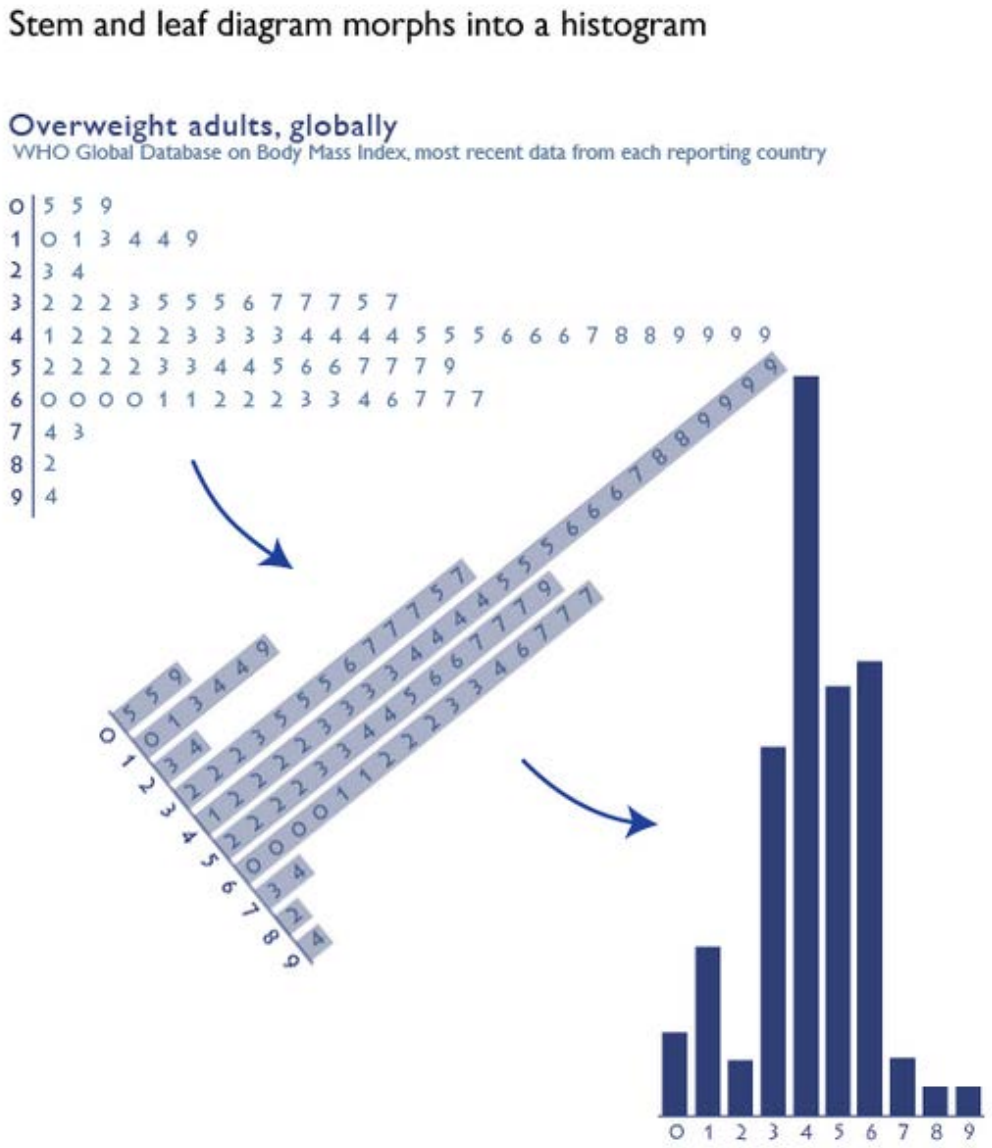
- Table
- Dot Plot
- Stem-Leaf
- Box Plot
- Histogram
- Curves



Khan Academy resource:
[Comparing Data Displays]

Image Source: Learn Zillion

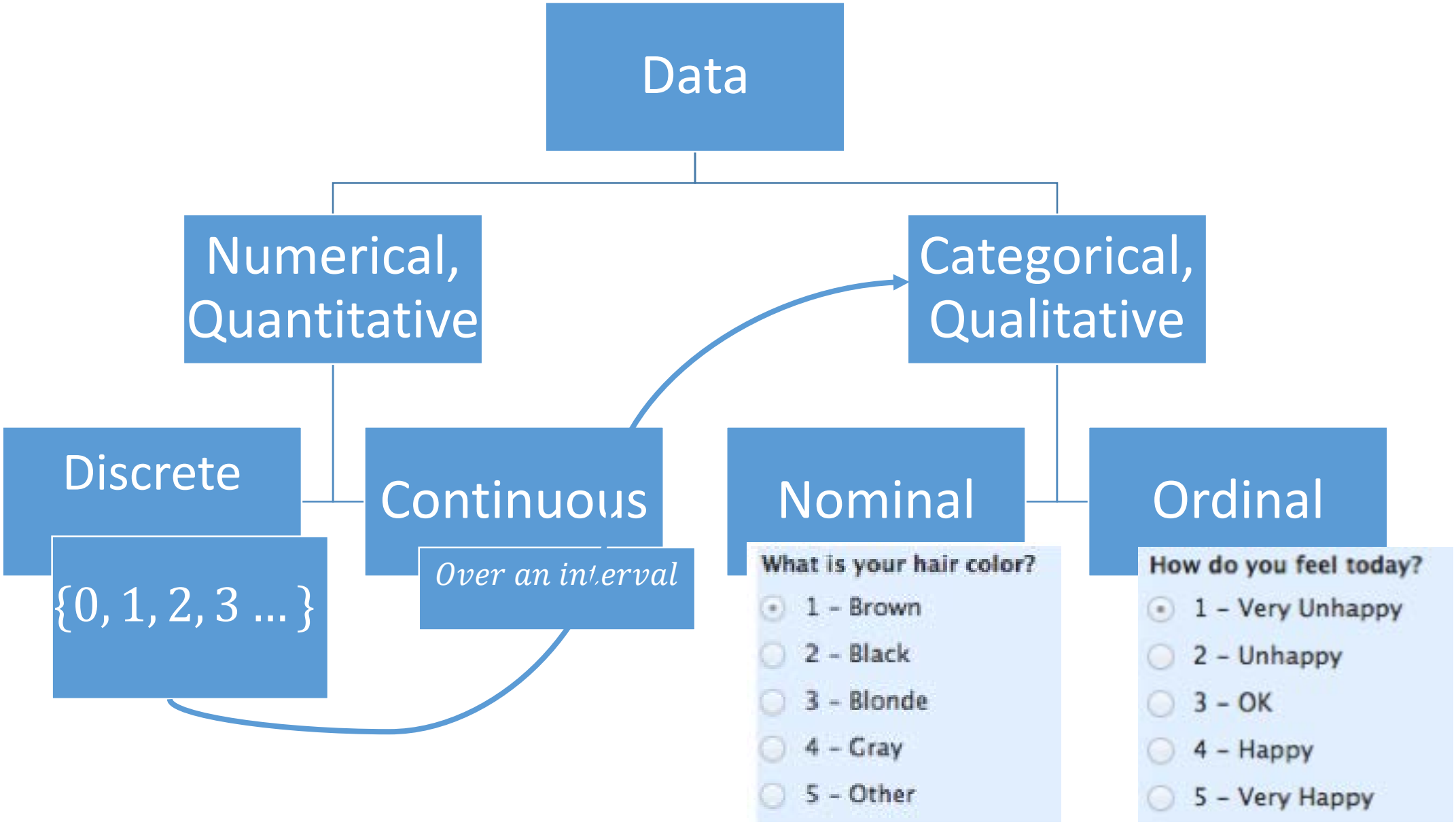
Stem and Leaf morphs into a Histogram



Stem-Leaf

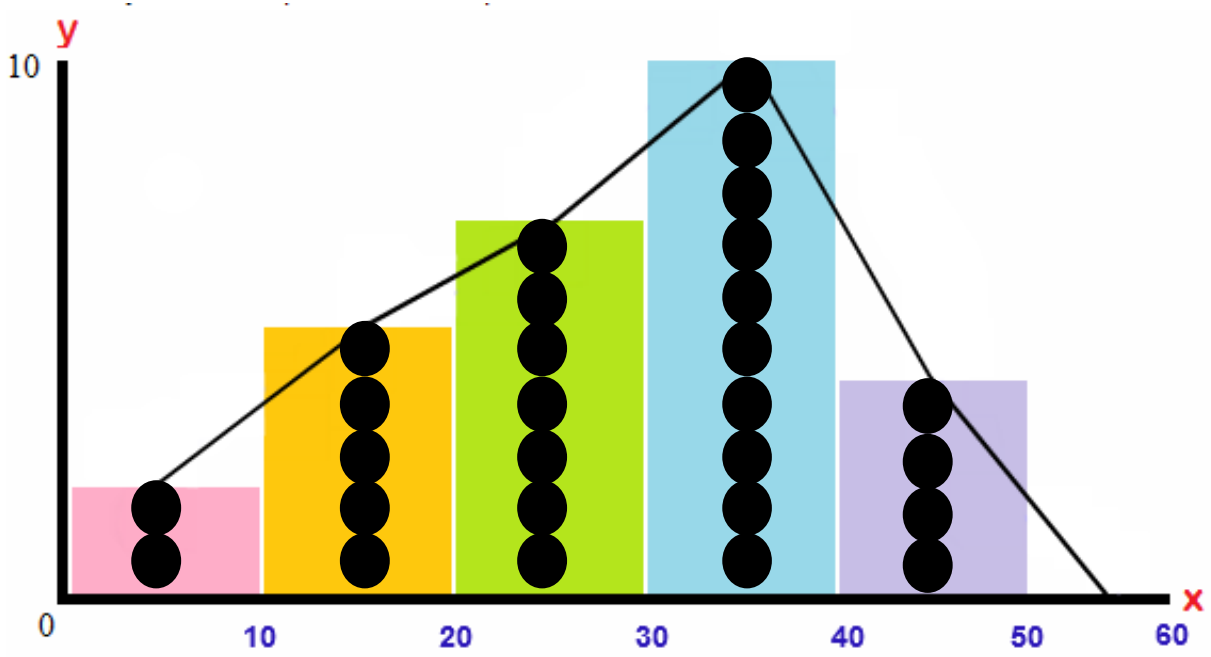
Stem	Leaf
22	8
23	1, 2, 5, 6, 7
24	3, 4, 7, 9, 9
25	0, 3, 5, 8, 8, 9
26	2, 2, 3, 5, 9

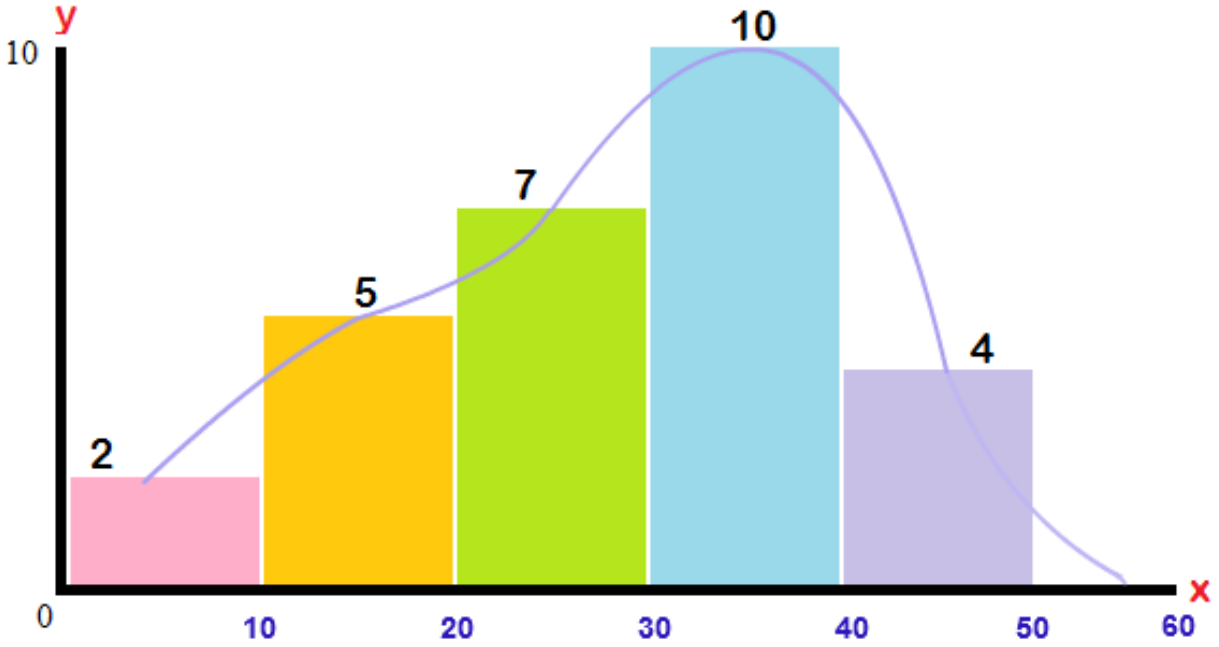
Key 22|8 is SAT score 228


Data
Numerical, Quantitative
Categorical, Qualitative
Discrete
{0, 1, 2, 3 … }
Continuous
Over an interval
Nominal
What is your hair color?
1 – Brown
2 – Black
3 – Blonde
4 – Gray
5 – Other
Ordinal
How do you feel today?
1 – Very Unhappy
2 – Unhappy
3 – OK
4 – Happy
5 – Very Happy

Distribution of Data

Dot Plots → Histograms → Frequency Polygons → Curves





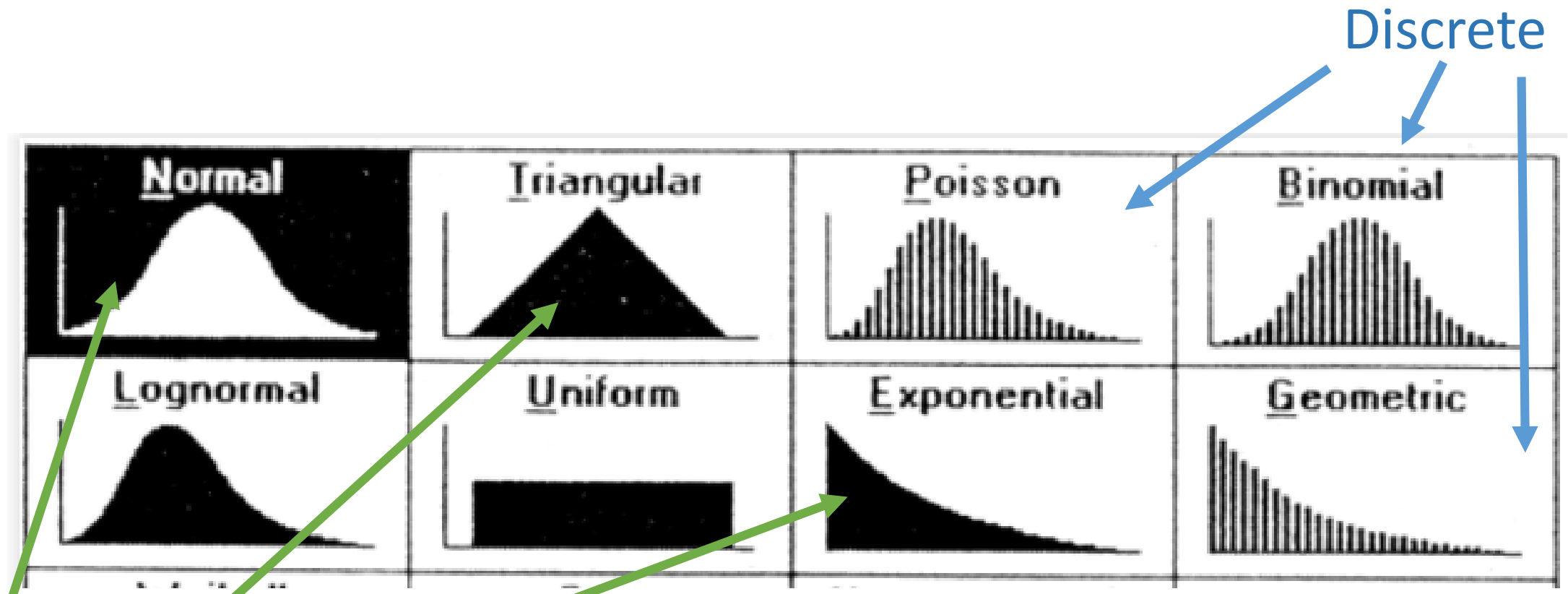
Where is the bulk of the data? (Funnel!)

What to you notice about the graph? (Focus!)

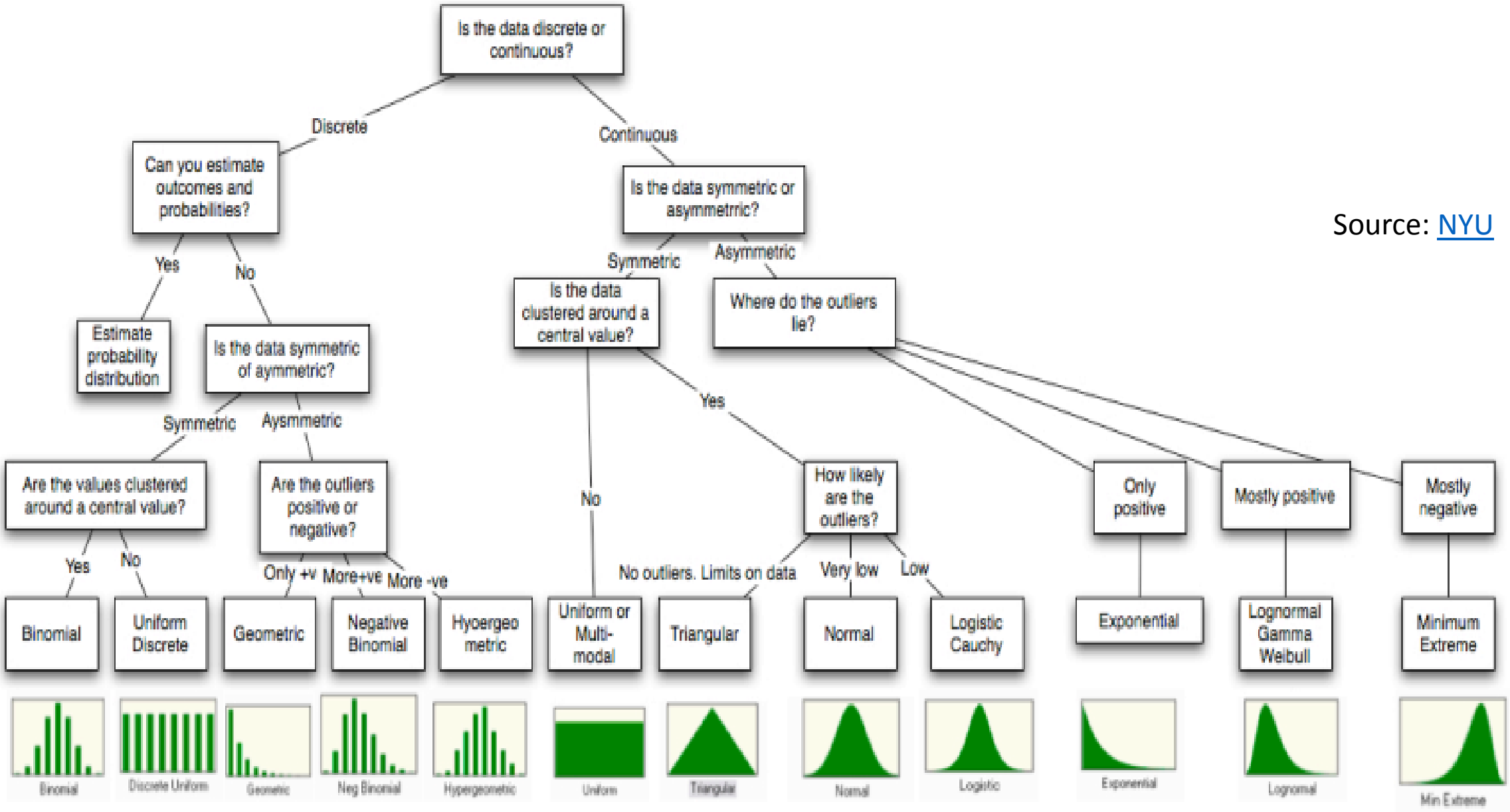
What percent of the data is less than 30?

What is the probability that a randomly selected unit is less than 30, $P(X < 30)$?

Common Distributions, Probability Density Functions (pdf's)



Modeling Data with Distributions*



Source: NYU

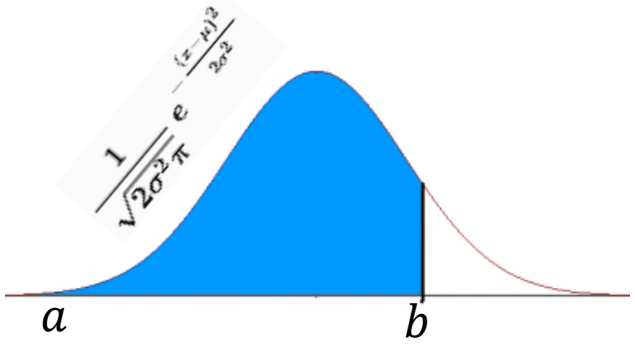
Probability Distributions

- Each Distribution (histogram or curve) is defined by a function, a model
- Discrete Data, take the sum of the model: $\sum_{i=1}^{n} p_i \cdot x_i$
- Continuous, integrate the model: $\int_a^b f(x)\,dx$

- Example Distribution: $N\sim(\mu, \sigma)$
  - Parameters mean, μ, and standard deviation, σ.
- Equation for the normal curve: $f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Its continuous, so integrate to find the area under the curve,

$$\int_a^b \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

- The area under the curve represents the probability for that event



Relative Frequency

- $y - axis$ represents percent or proportion instead of actual count
- Same shape as Frequency Histograms

Cumulative

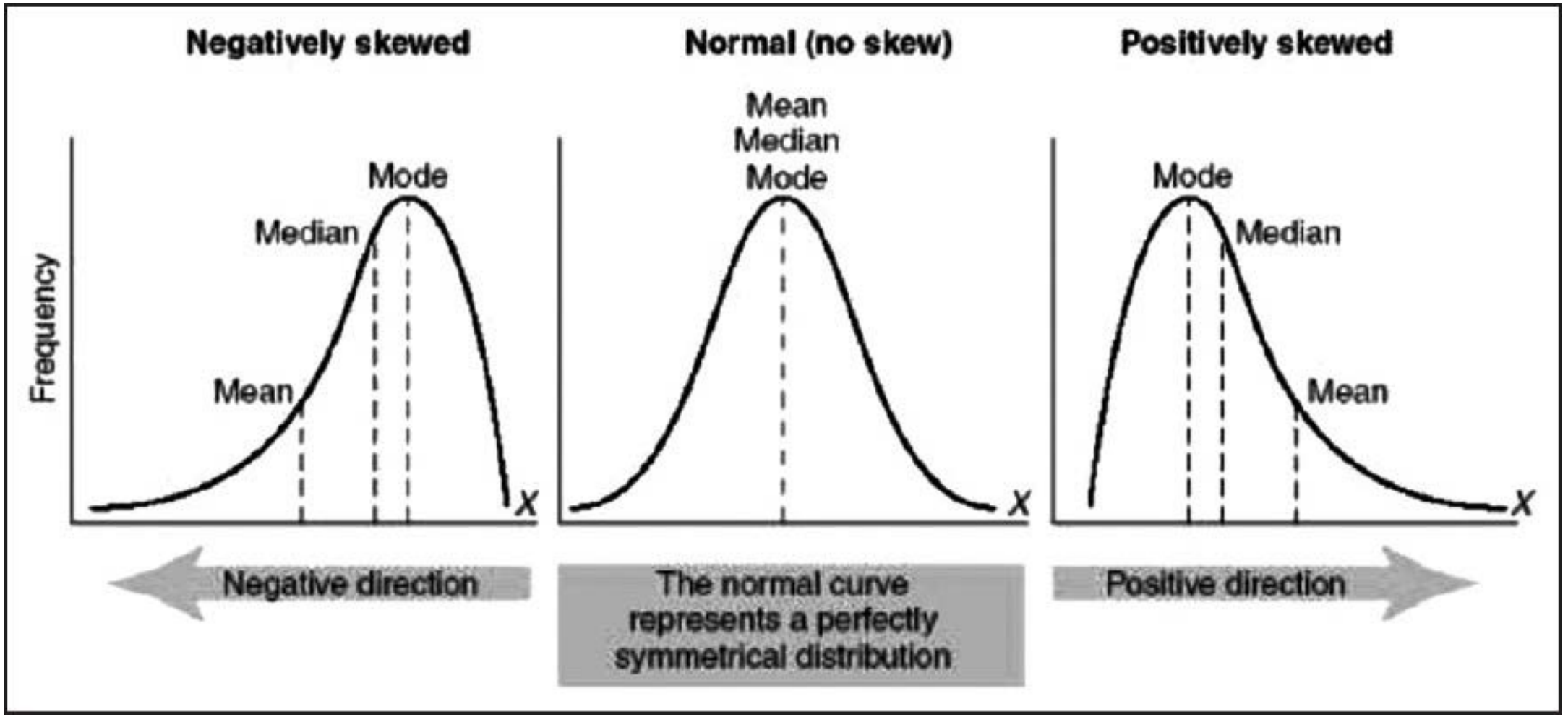
Pareto

Describe how the data is distributed (SOCS)

- **S**hape
- **O**utliers
  - An observation that outside the pattern or distribution
- **C**enter
- **S**pread

Source: Mathworld Wolfram

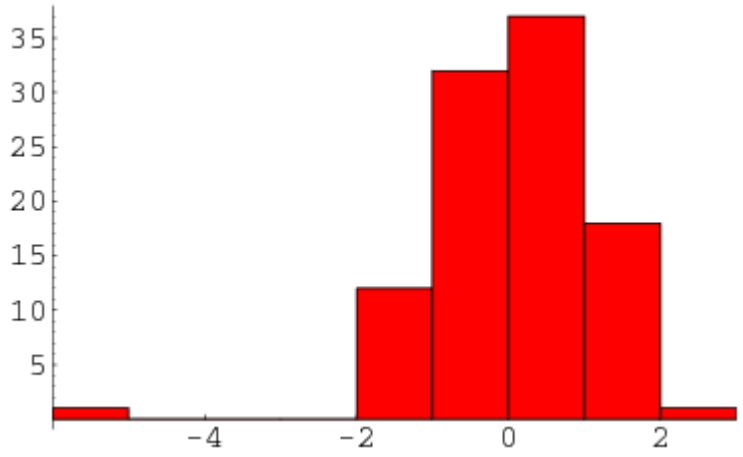
Shape



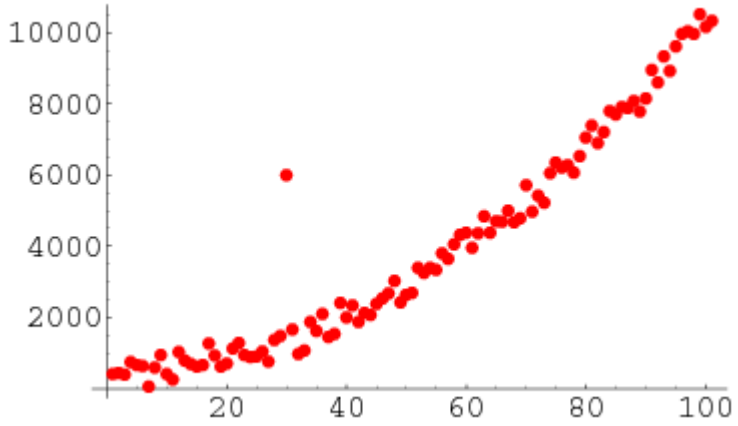
Source: **Descriptive statistics. Available from** http://www.southalabama.edu

Identifying Outliers

Histogram



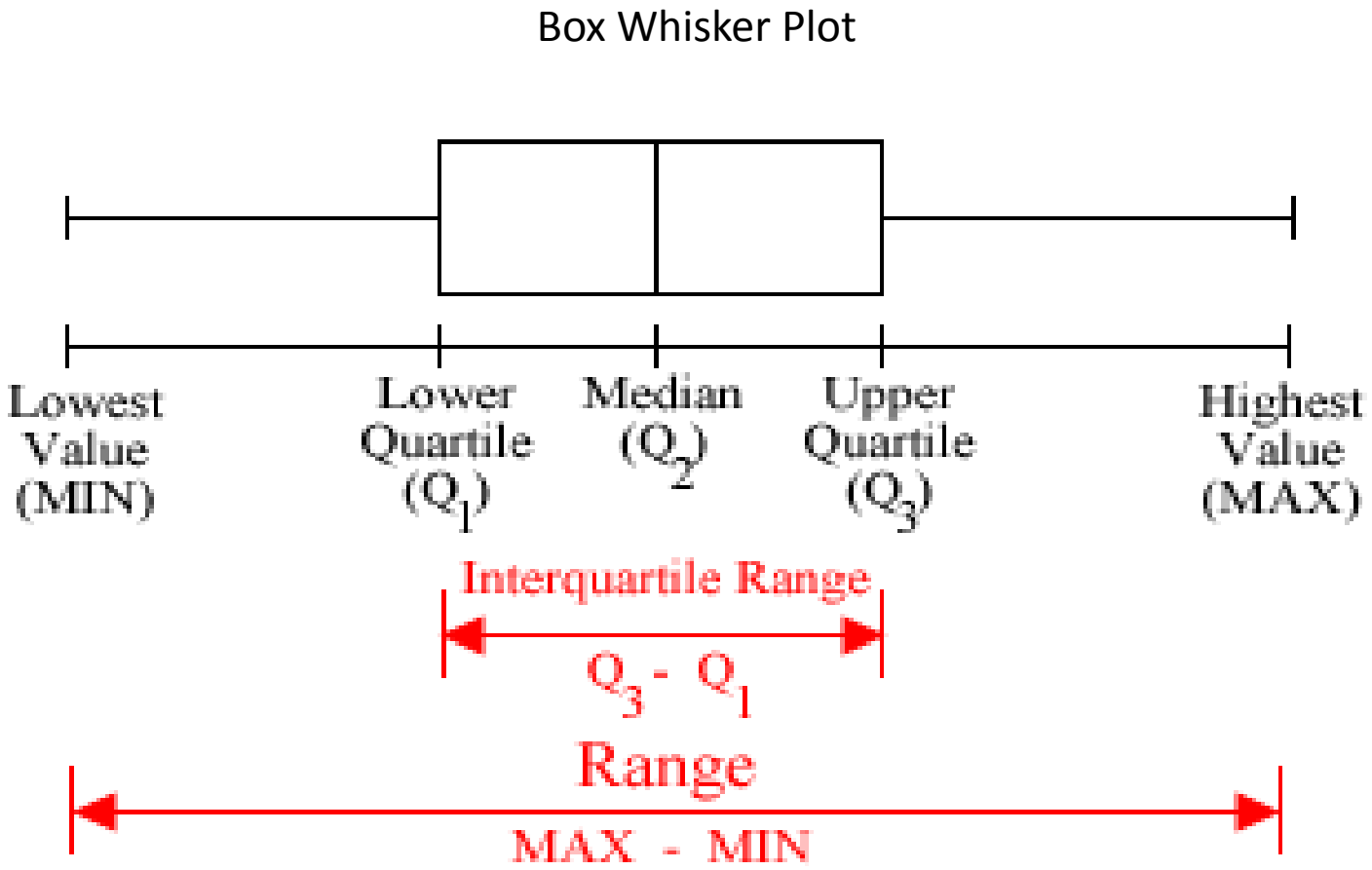
Linear Regression



Source: [Mathworld Wolfram]

Computing Outliers

- $Q1 - 1.5(IQR)$
- $Q3 + 1.5(IQR)$



Descriptive Statistics

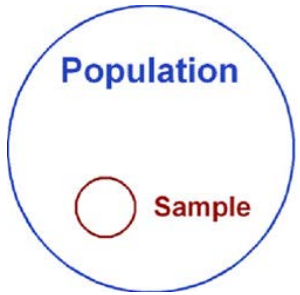
Center

- Mean
- Median
- Mode

Spread – Variability, Dispersion

- Standard Deviation
- Variance
- Range
- Interquartile Range

Statistics Notation



Data	Population	Sample
• Mean	• μ	• $\bar{x}$
• Standard Deviation	• σ	• s
• Number of data points	• N	• n
• Characteristic being measured	• Parameter	• Statistic

Formula for Mean

Same Formula

Sample Mean	Population Mean
$\bar{x} = \frac{\Sigma x}{n}$	$\mu = \frac{\Sigma x}{N}$

where Σx is sum of all data values

N is number of data items in population

n is number of data items in sample

Formulas for Standard Deviation

Different Formula

Sample

- $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}}$

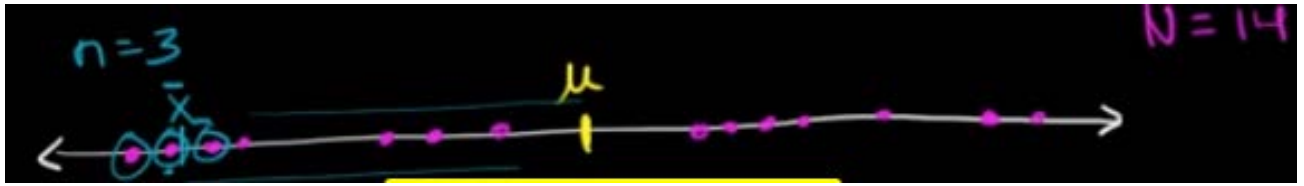
Population

- $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)}{N}}$

- Variance
  - $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)}{N}$

Why is the Standard Deviation of the Sample divided by (n-1)?*

- Dividing (n-1) instead of n
  - Gives an unbiased, more accurate estimate, of the population.
  - A larger value
- If you divide by n,
  - Then you will get a smaller value and likely to underestimate the standard deviation, distance from the mean.



Note: Population Mean and Sample Mean are the same formula because the Sample Mean will always be within your sample.

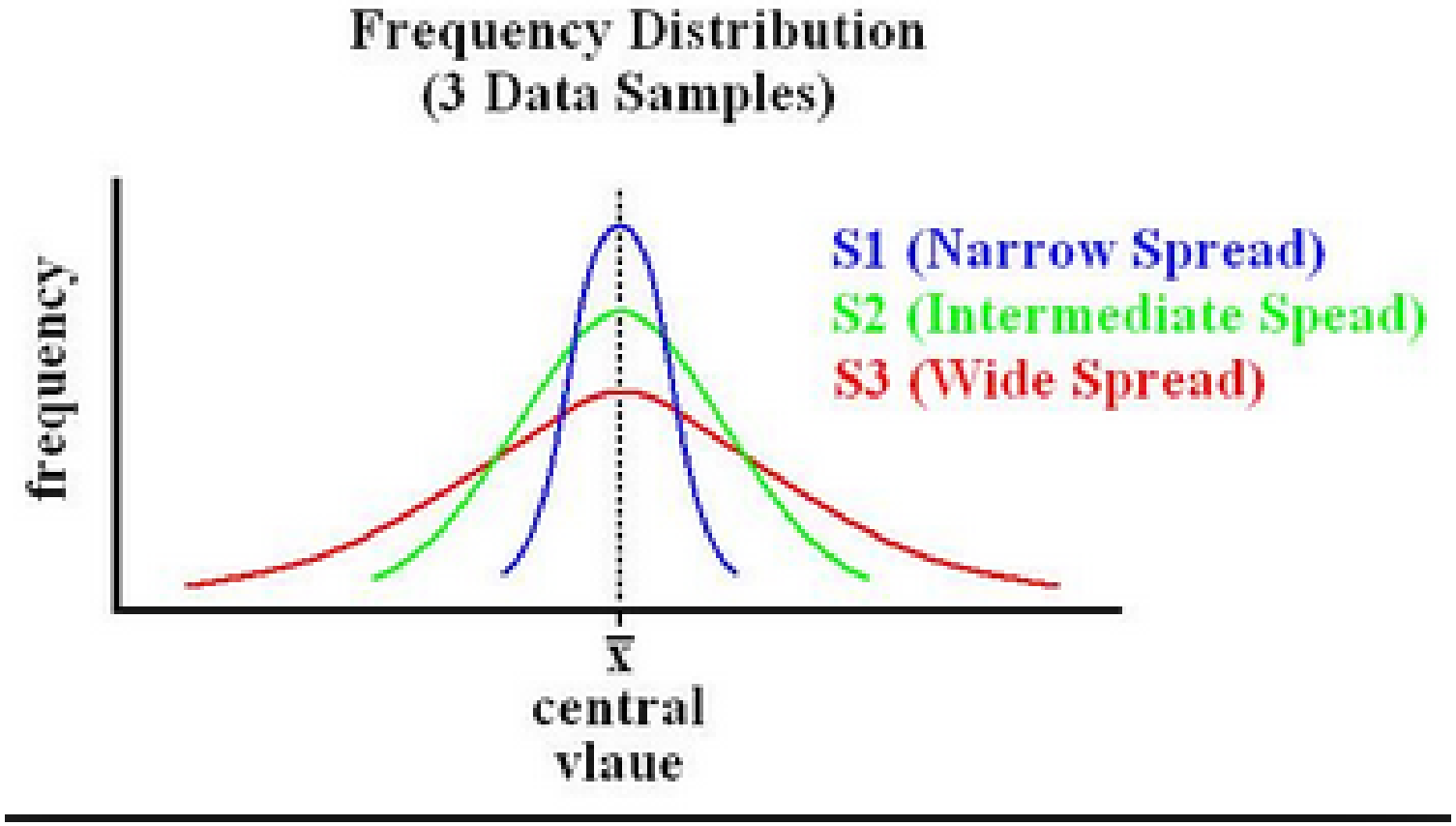
Khan Academy, why (n-1)?

Khan Academy, Simulation for more intuition on (n-1).

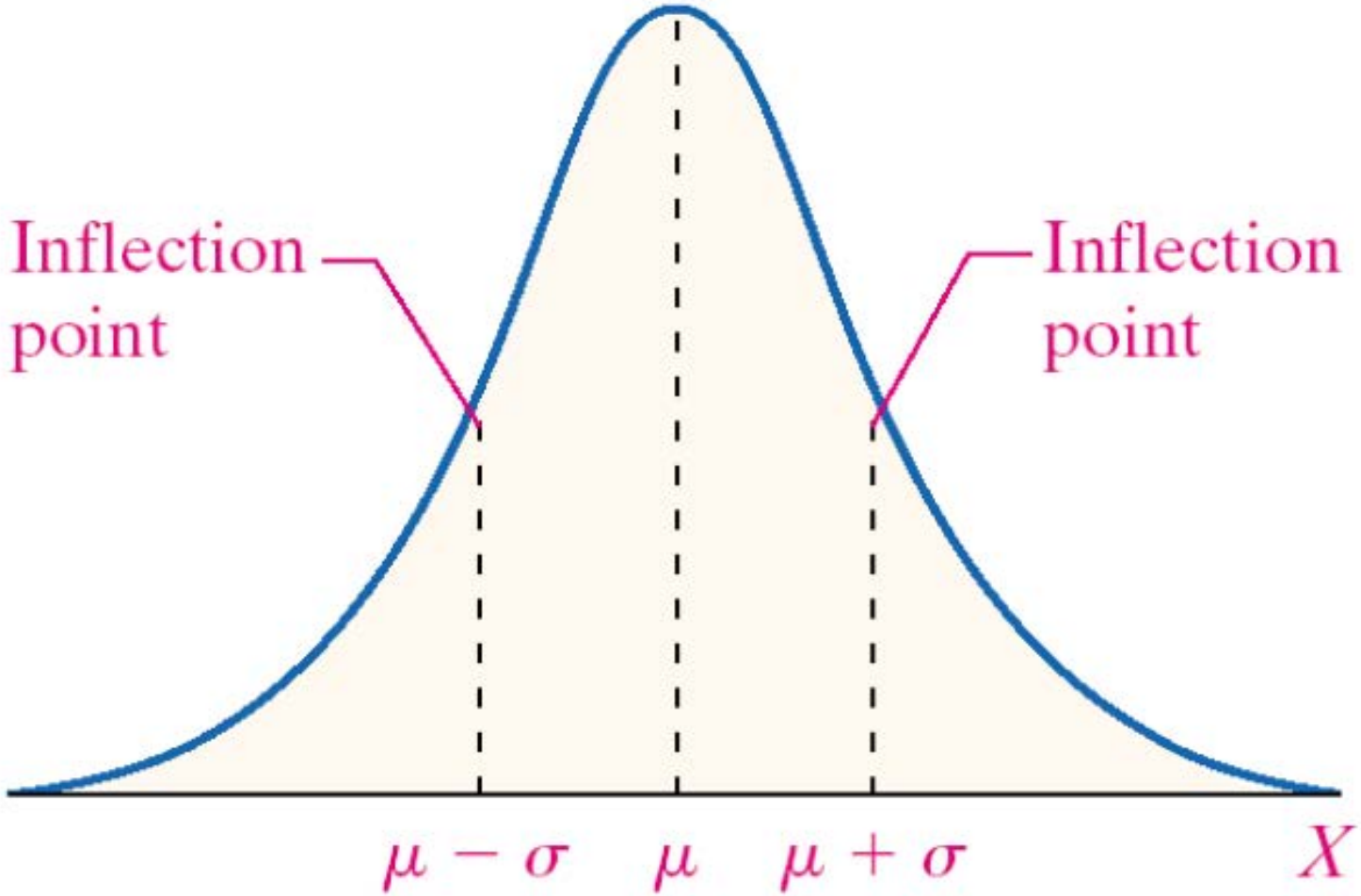
Intuition behind Standard Deviation, Variability
(Variation or Spread)

- Make a Dot Plot of Shoe Size on the board
- Unimodal or Bimodal?
- Narrow data to women's shoe size
- Compute the sample mean, $\bar{x}$, and label on dot plot
- Draw the distances from each data point to the mean
- Square each distance
- Compute the average of the distances
- Take the square root

SPREAD-Variability



Estimating Standard Deviation from a Graph



Image

Computing Standard Deviation by Hand

x	$(x - \bar{x})$	$(x - \bar{x})^2$
$n =$ $\bar{x} =$		$\sum (x - \bar{x})^2 =$

$$variance = \sigma^2$$

$$= \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$standard\ deviation = \sigma$$

$$= \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Question 1. How many years does a penny stay in circulation?

- Make a list of ideas to go about how to answer this question.
- What is a good sample size?
- How you gather the pennies for the study?
- See Representing Data Worksheet.

Representing Data

Rep. Data Penny

Rep. Data Penny Tech

How many years does a penny stay in circulation? (Teaching Notes)

- Funnel and Focus Questions
- STEM approach
- Data
  - Dot Plot
  - Histogram, Frequency Chart
    - Put borderline data values in right bin
  - Relative Frequency Chart
  - Box Plot on TInspire

Part 2: Experimental Design

How do you know if claims in news headlines are true?

What would discredit a study or poll?

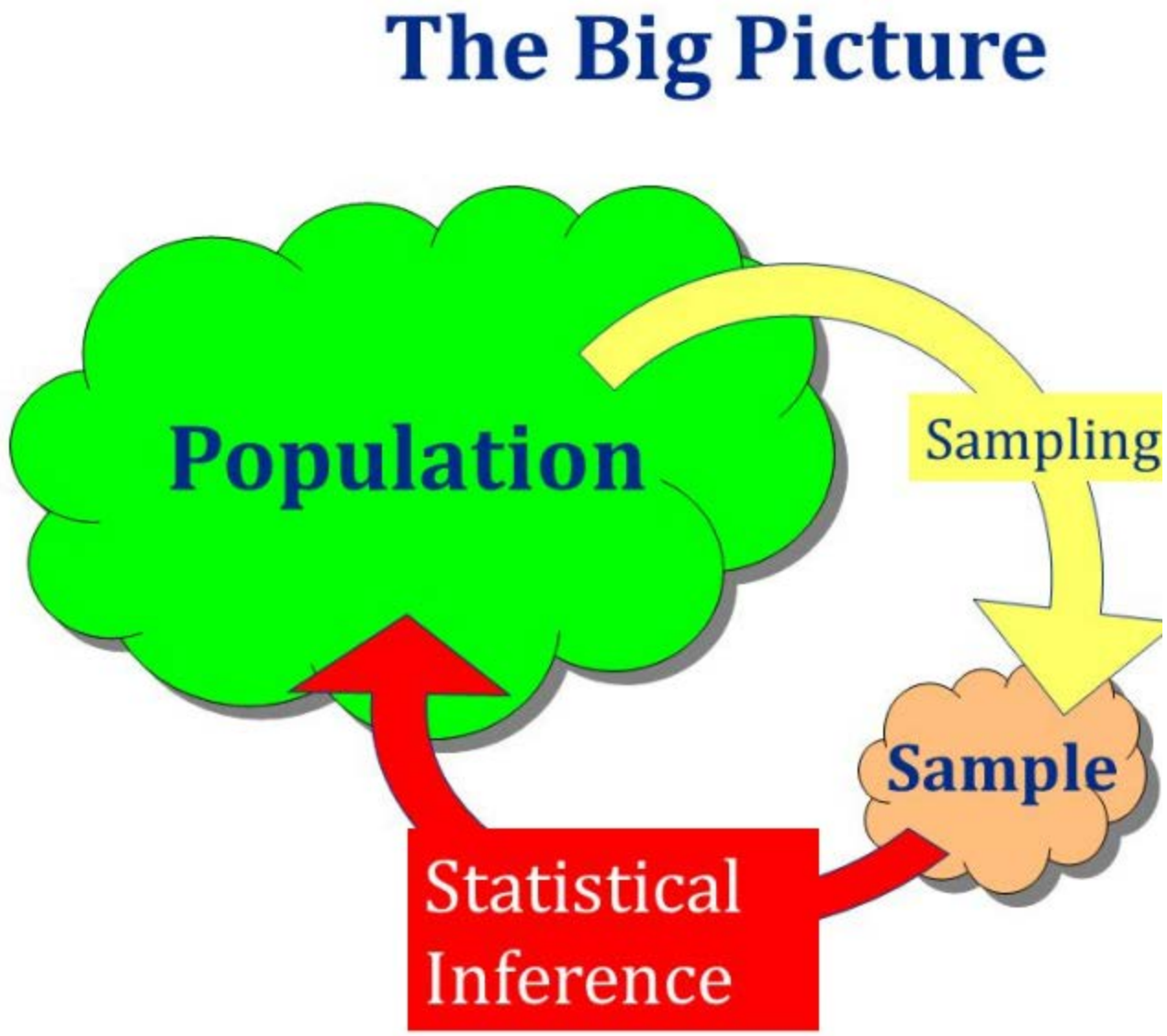


Image Source: SlidePlayer

Question 2: What is the “readability” of the *Gettysburg Address*?

- Make a list of ideas to go about how to answer this question

Gettysburg

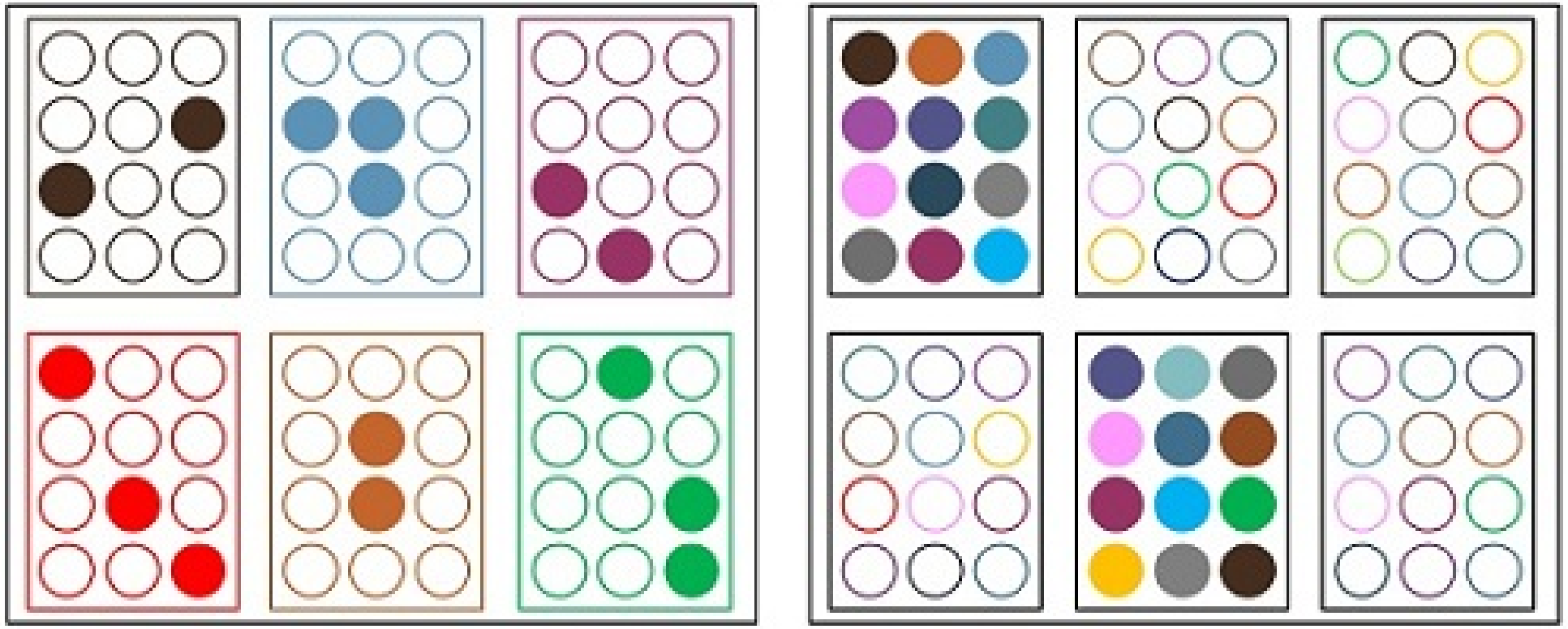
Gettysburg
Google Tech

Detect the bias in the sampling and surveying examples and explain why it is biased.

- Standing outside McDonald's asking every third person, how many servings of vegetables the ate today?
- A researcher nodding and smiling while asking a patient if they feel better after treatment?
- An MSNBC internet poll
- Asking, do you enjoy drinking alcohol?
- Do teenage moms deserve government paid daycare?
- What was your ACT score?
- A University requesting alumni to email back their job status.

How to Reduce Sampling Bias?

- Simple Random Sampling (SRS)
  - Each element has equal chance
  - Even the most advanced computer random number generator is not a pure random number
- Unbiased Survey Questions
  - Neutral, not leading or loaded questions
- Systematic Random Sample
  - Every 10^{th} person
- Convenience Sampling
  - Convenient sample because of accessibility or proximity
- Voluntary Response Sample
  - Self selected sample
- Response Bias
  - Subject answer untruthfully or misleading
- Stratified Random Sample
  - Randomly selecting from each homogenous group
    - Example: Select a few families from each group, i.e. Urban, Suburban, and Rural settings
  - Ensures every part of the population is represented
  - Difficult when population is large
- Cluster Sample
  - SRS several clusters, assuming each cluster is representative of the population
  - Example: Assume Urban, Suburban, and Rural families are represented in each cluster, then selecting a few clusters
  - More convenient than Stratified, reduces cost while increasing efficiency



Stratified Sampling Vs Cluster Sampling

Image Source

Be a Skeptic!

- Who was the author of the study?
- Who published the study?
- Were opposing views represented?
- How big was the sample?
- How was the sample obtained?
- Let me see the data.
- What was the hypothesis?
- The U.S. National Science Foundation defines three types of research misconduct: fabrication, falsification, and plagiarism, (Wiki, 2016).

Gold Standards of Experimental Design

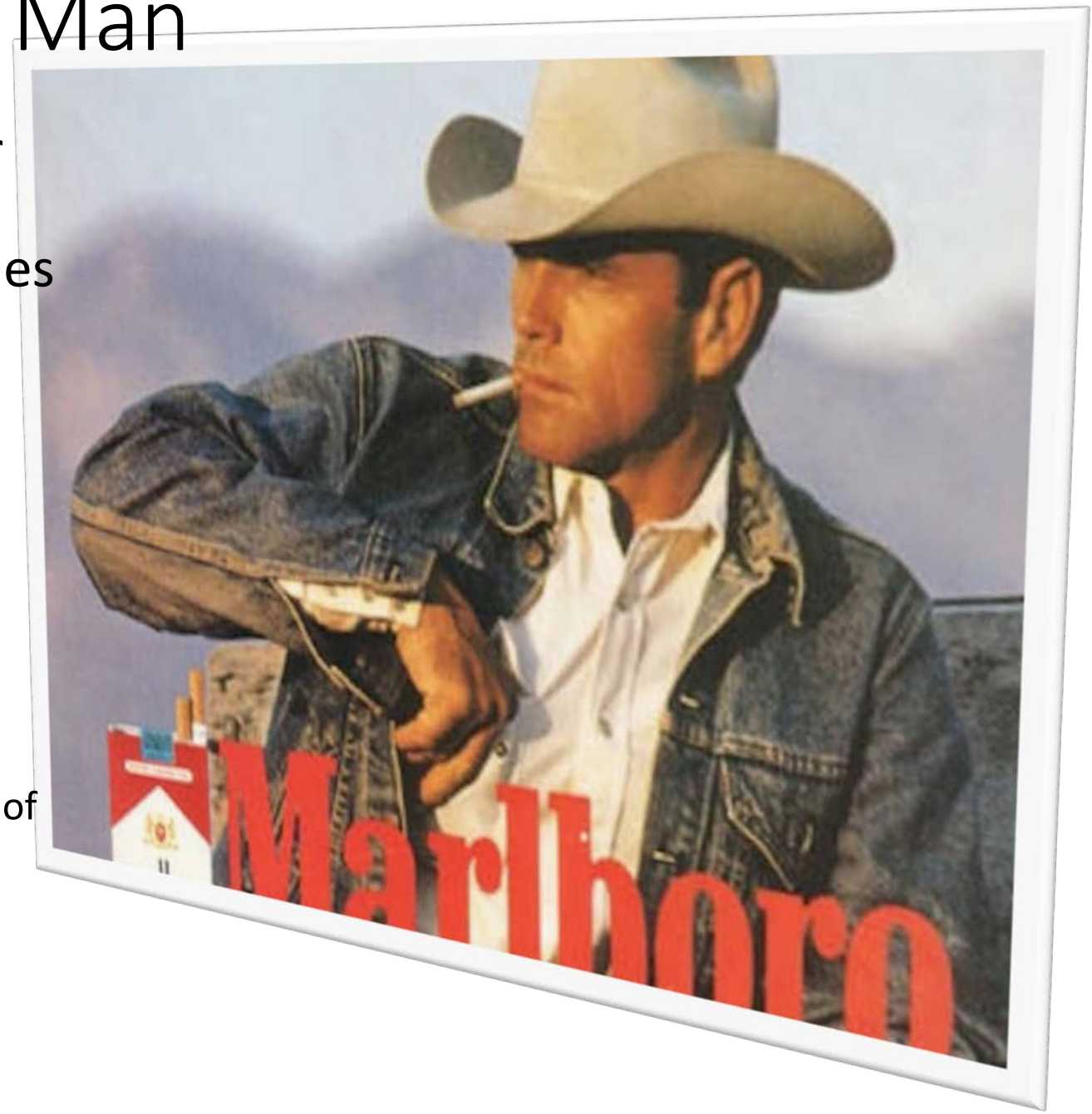
- Sample is
  - Random
  - Controlled
  - Large
- Double Blind
  - Neither researchers or subjects know which subjects are in the experiment or control groups
- Placebo

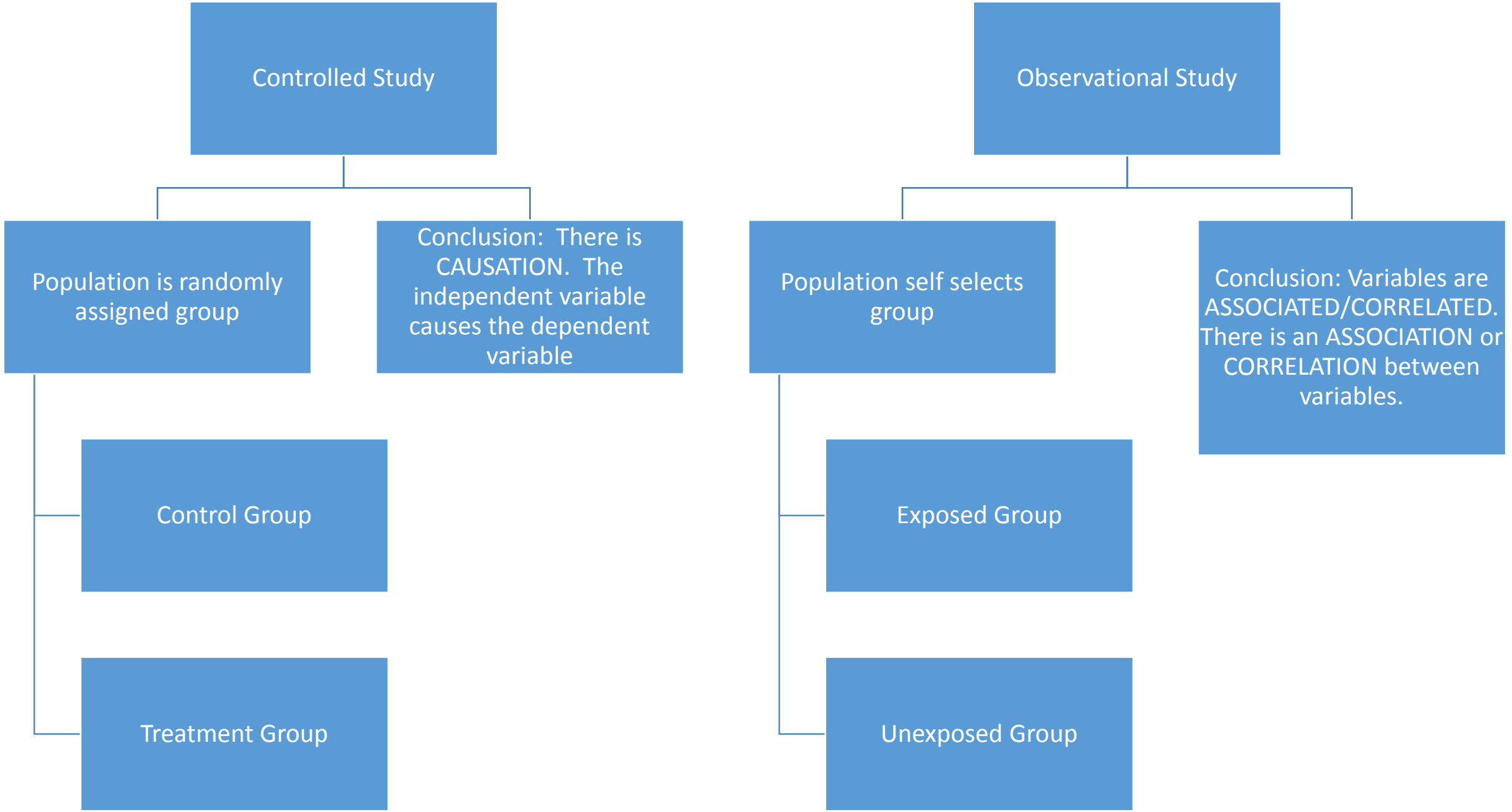
What does Controlled mean?

- Experimental (Treatment) and Control Group (Baseline)
- Groups must be as close to identical as possible, very similar
  - Otherwise the treatment being tested could be explained by something else, such as a confounding variable.
- Tests only one variable

Marlboro Man

- "Smoking Causes Lung Cancer," Phillip Morris Co. finally admitted in 1989 after decades of denial! (NYT, 1999).
- Tobacco Industry and independent studies showed smoking caused lung cancer as early as 1946.
- How did the tobacco industry dispute these claims for 50 years?
  - Argued the control and treatment groups were not similar
    - Could be explained by sex, genetics, air pollution, viruses, etc.
  - It was an observational study, not a randomized controlled study
    - A randomized controlled study on the effects of smoking would be unethical
- A study of identical twins was the main factor that led to the tobacco industry admit their product caused lung cancer.




Controlled Study
Population is randomly assigned group
Conclusion: There is CAUSATION. The independent variable causes the dependent variable
Control Group
Treatment Group
Observational Study
Population self selects group
Conclusion: Variables are ASSOCIATED/CORRELATED. There is an ASSOCIATION or CORRELATION between variables.
Exposed Group
Unexposed Group

Facebook Case-Violation of Informed Consent?

- In 2014, FB did a controlled randomized study on nearly 700,000 Facebook users without informing them.
- One group viewed more positive posts and one group viewed a more negative feed to see if it influenced them.
- What's the big deal? No one was hurt!
- But, it was harmful because it was unethical and broke the law.
- Facebook may have had permission to collect data
- But, they did not disclose that users were participating in a study and what the study was about. The Facebook users were not informed!
- Also, it is possible someone did get hurt. Of those 700,000 users, maybe just one viewing the negative news feed was prone to depression or suicide and the negative news feed was too much to handle.
- Even if there was no physical harm from the study, Facebook attempted to influence the minds of 700,000 people. That is dangerous because it is so powerful. And, it turns out, the Facebook's study showed positive and negative news feeds influenced the users posted accordingly.

Tuskegee Syphilis Case

- Wiki Tuskegee
- Informed consent, a human right
- US Public Health Service wanted to study the natural progression of syphilis
- Untreated and lied to rural African American for 40 years
- Untreated subjects, their spouses, and offspring were harmed which included death.

Variables-A quantity or quality that varies

- Independent
  - Believed to affect dependent variable
- Dependent
  - Researcher is interested in
- Extraneous
  - Little to no influence on experiment
  - Any variable other than the dependent and independent variable
- Confounding
  - Interfere with study
- Control
- Kept the same in each trial
- Moderator
  - Increase or decrease relationship between independent and dependent variables

Q3: Who is the most elite athlete?

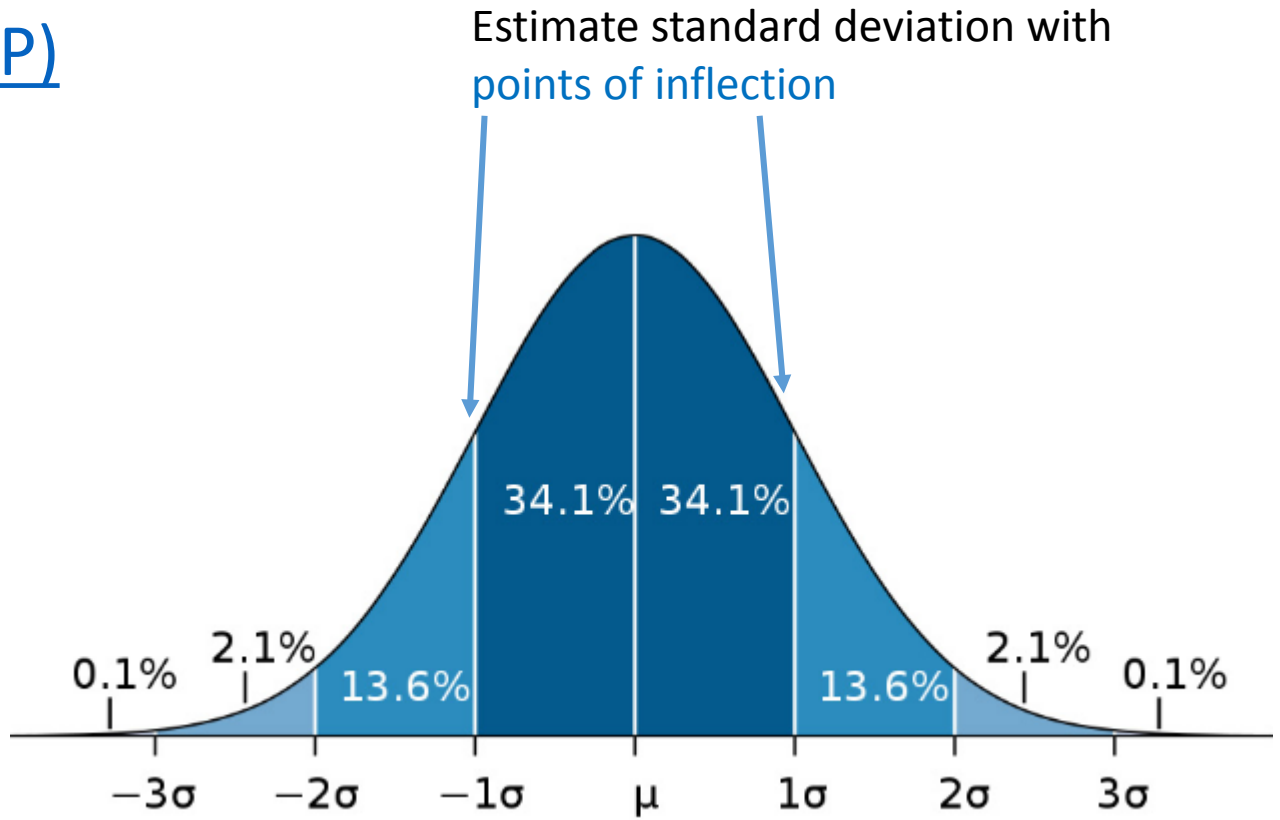
- Cross Country Runner
- Swimmer
- Bicyclist

Elite Athlete

[Iron Man Stats]

Part 3: What is Normal?

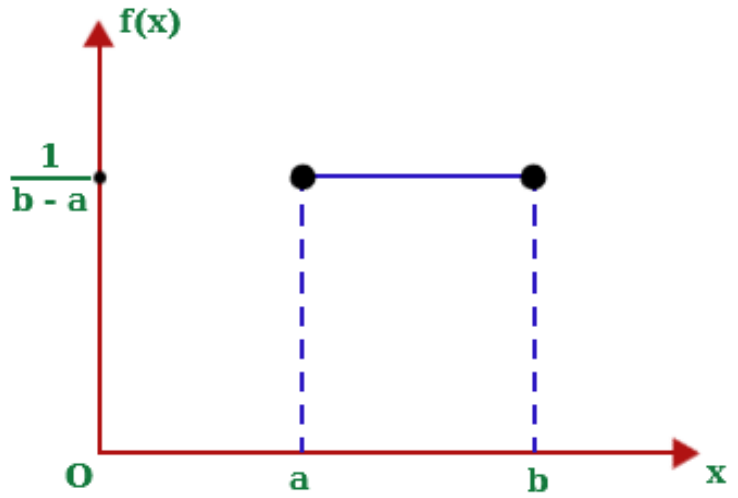
- Normal Distribution $N \sim (\mu, \sigma)$
- Mathematics Vision Project (MVP)



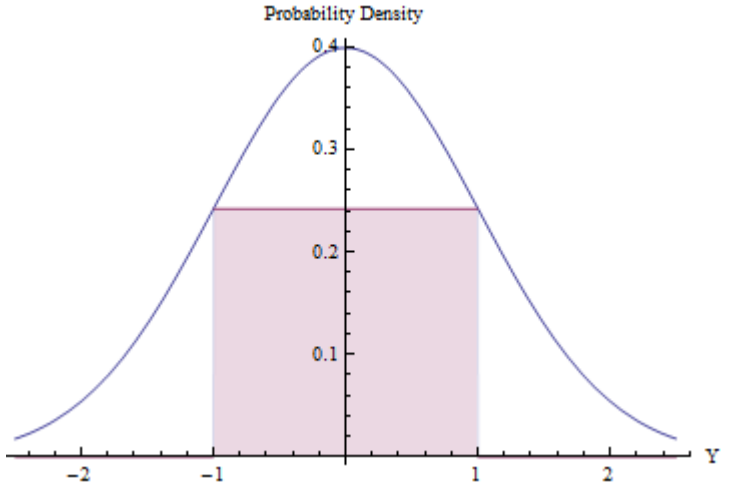
- Properties
  - Unimodal
  - Symmetric
  - Tails extend to infinity
  - Area under the curve, [0, 1]
  - Area represents the $Probability, 0 \leq p(x) \leq 100\%$
  - Determined by it's mean and standard deviation

Why is the Area under the curve 1?

- Area is 1 for all distributions
- Easy to see on a Uniform Distribution
  - Try values for a and b



- 100% of the events are represented under the curve



Estimate mean and standard deviation from graph.

$N_1 \sim (\quad , \quad)$

$N_2 \sim (\quad , \quad)$

$N_3 \sim (\quad , \quad)$

$N \sim (\quad , \quad)$

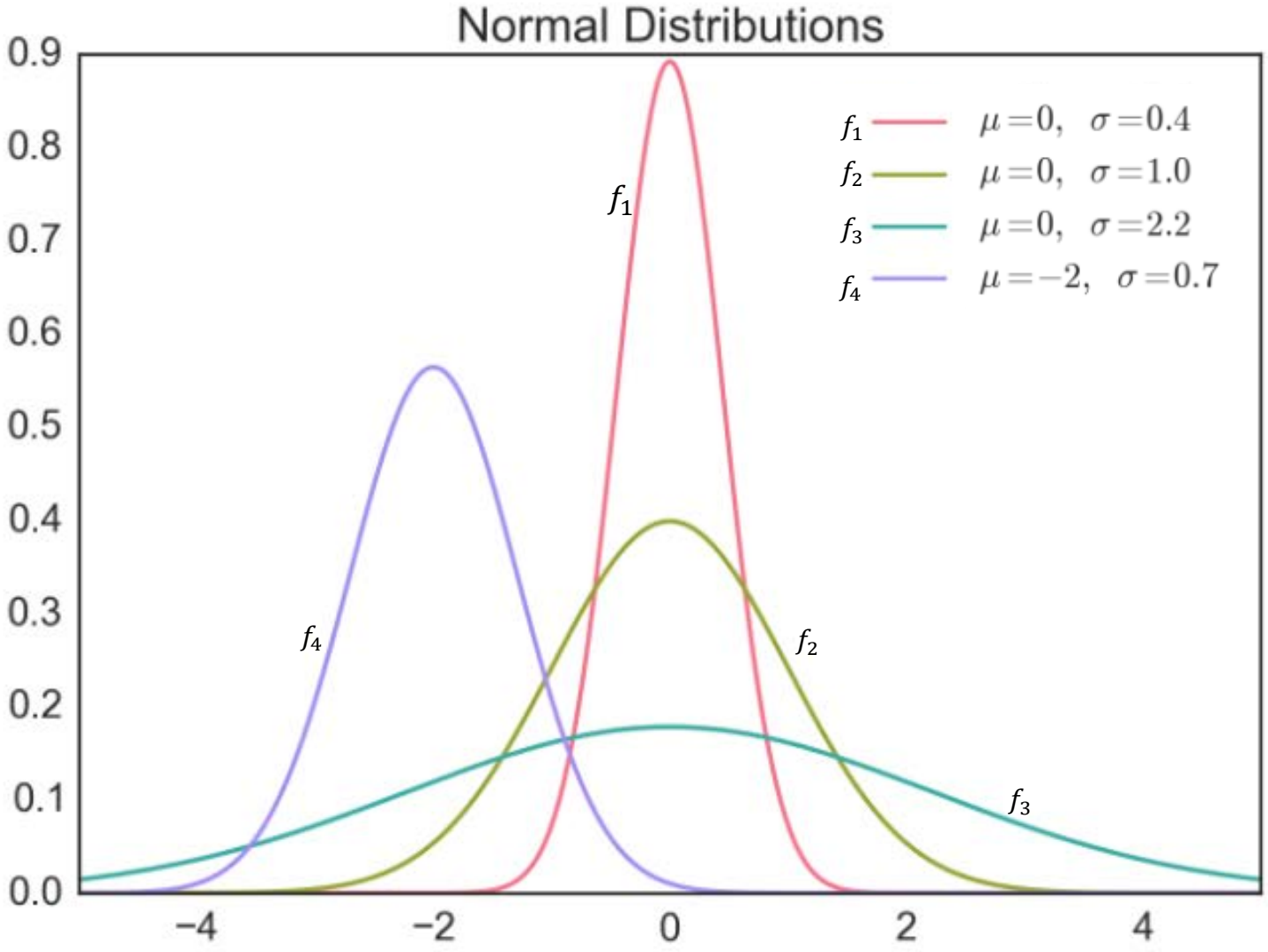


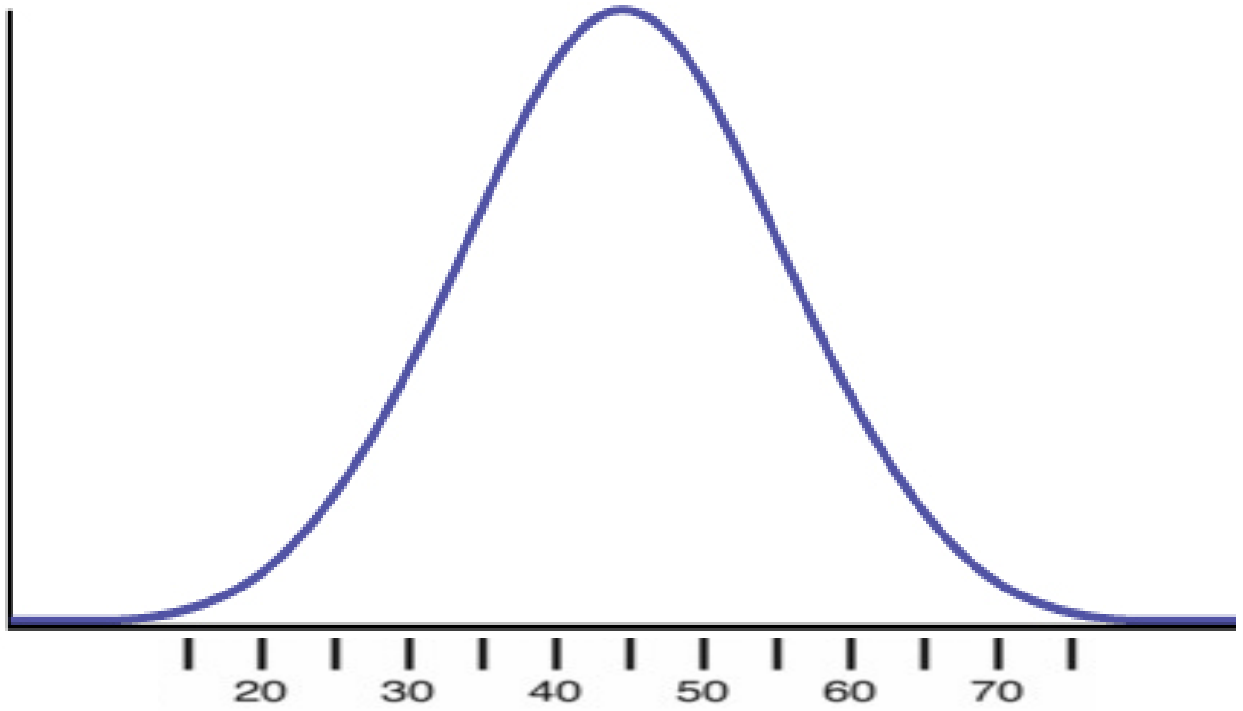
Image source

Challenge!

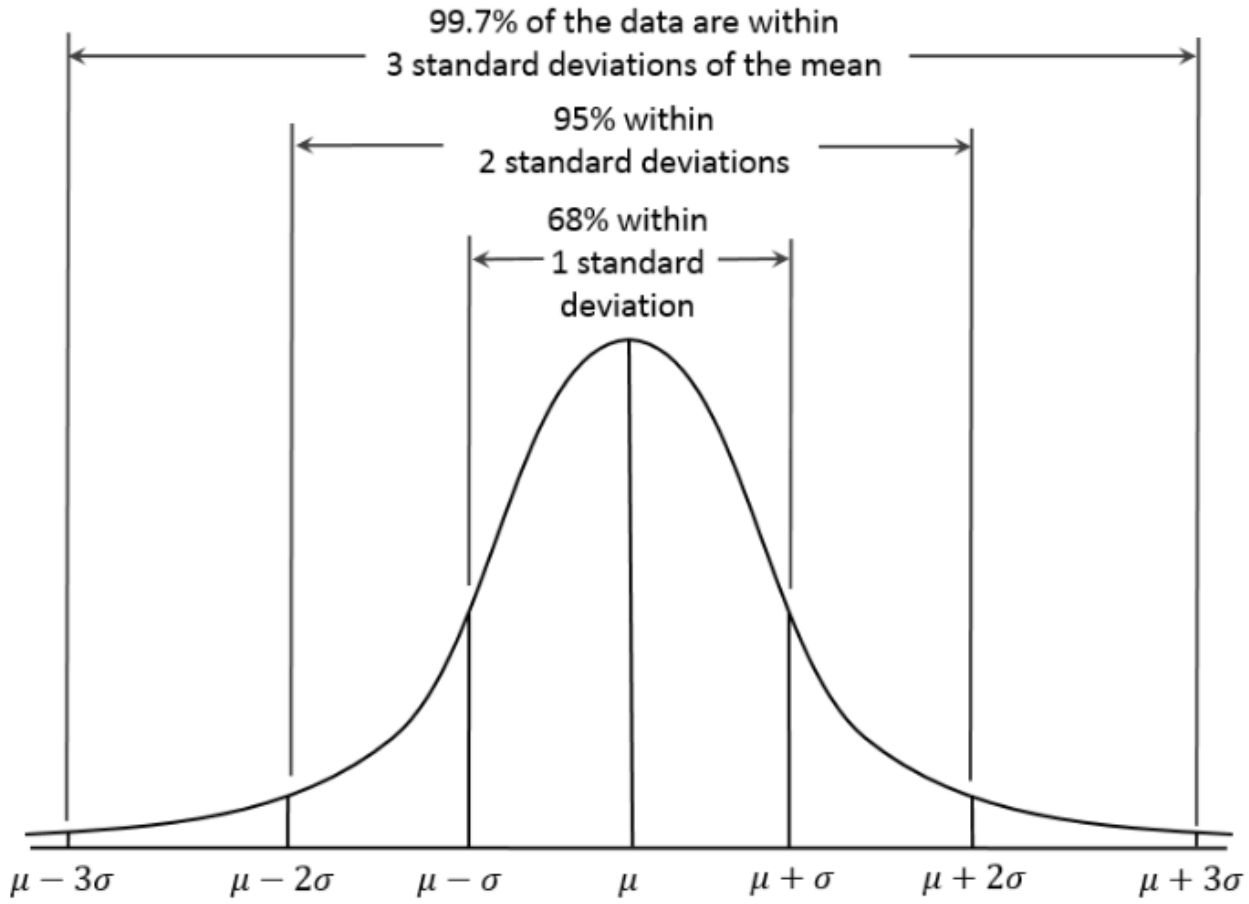
- Draw two normal curves, y_1 and y_2 such that $\mu_1 < \mu_2$ and $\sigma_1 > \sigma_2$.

Example: Estimating mean and standard deviation

- Find the average age and standard deviation of a Pearl Jam fan.



Empirical Rule



Image

Example: The average highway speed is given by $N \sim (60, 5)$. What % of drivers drive slower than 55 mph?

The Empirical Rule

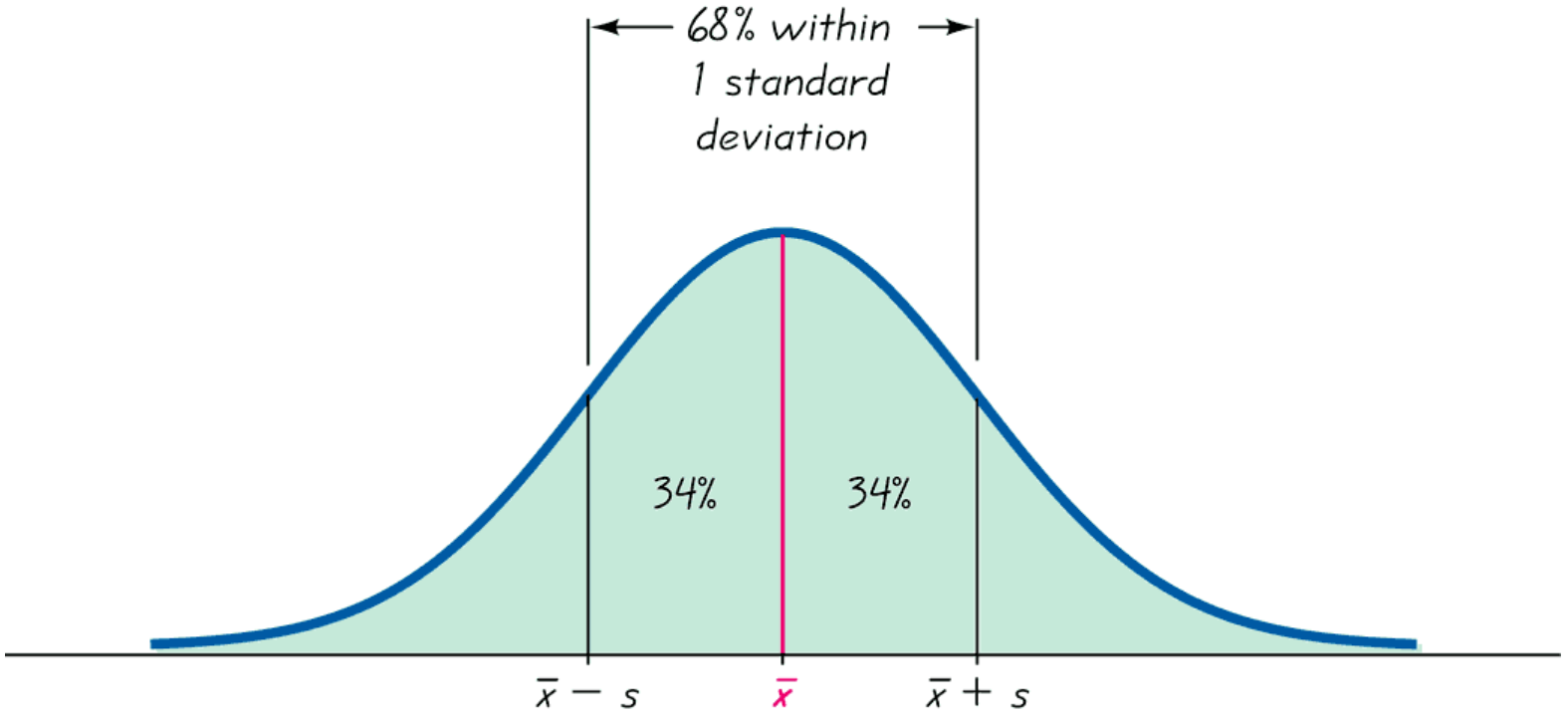


Image source:
Pearson, 2010

The Empirical Rule

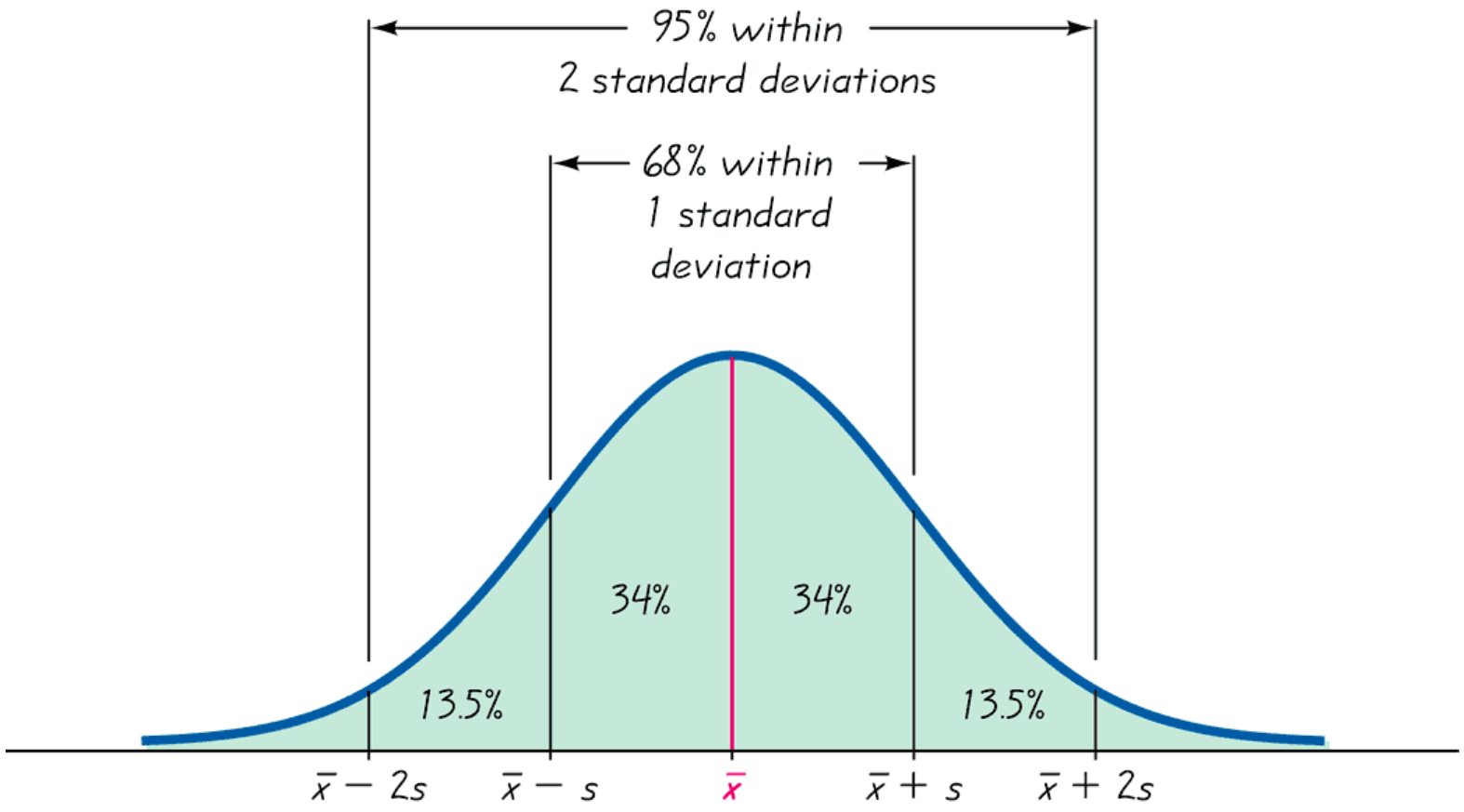


Image source:
Pearson, 2010

The Empirical Rule

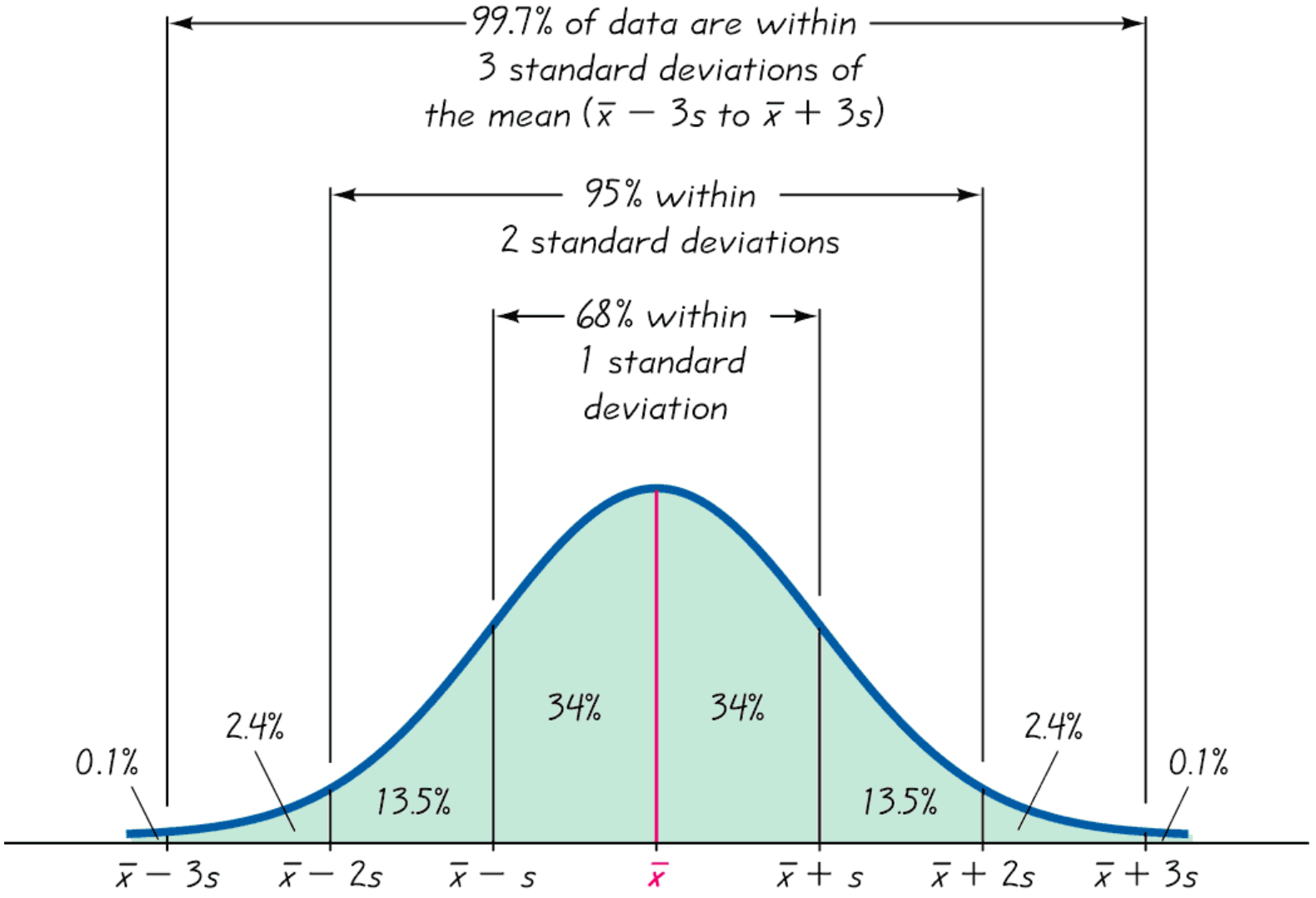


Image source:
Pearson, 2010

Compare Box Whisker Plot and Normal Distributions percentiles

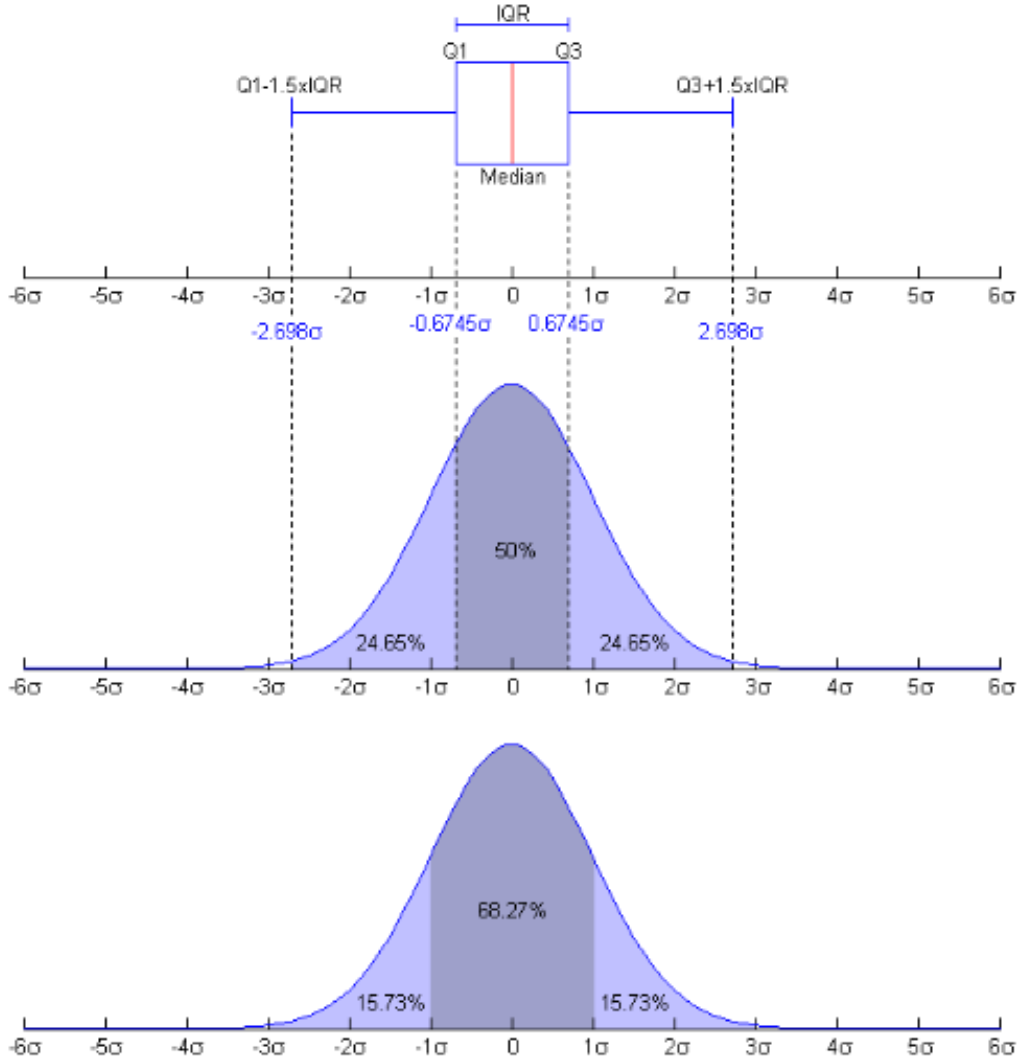
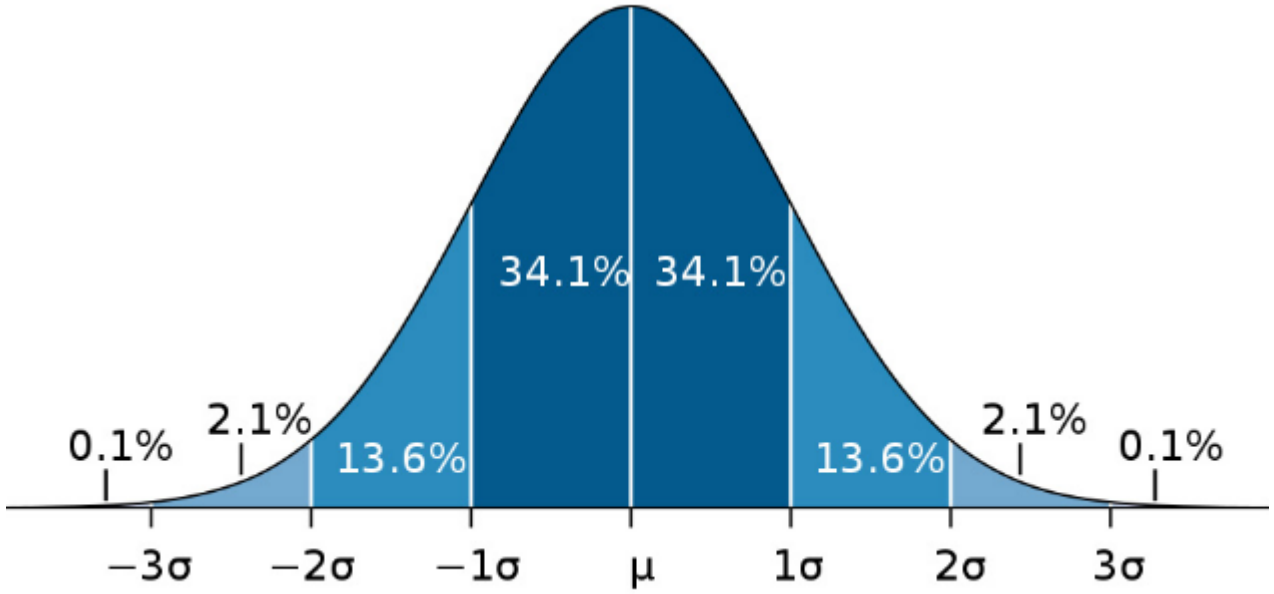


Image source: [Wiki]

$N \sim (60, 5)$

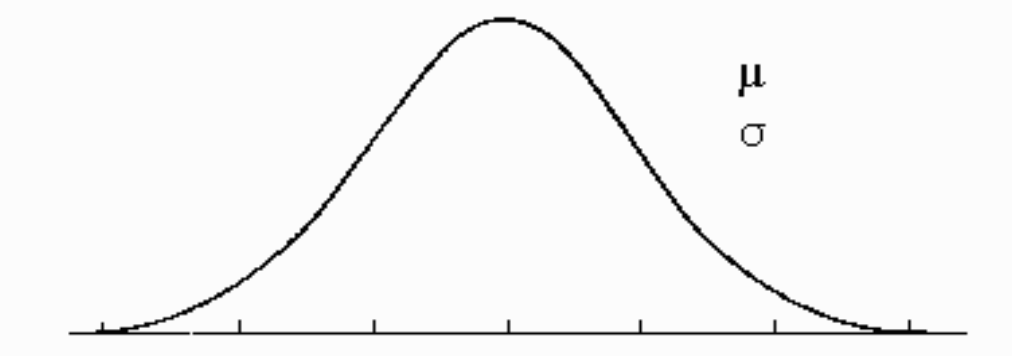


If twenty students were in the class, how many received a 65% or better?

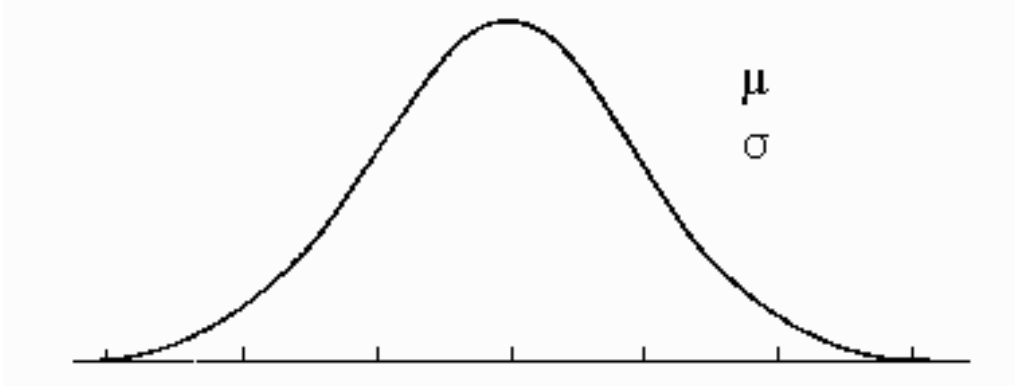
If a student was chosen at random, what is the probability he would have earned below a C? $P(X < 70) =$

Example: What is 92% a more impressive score for your math or history class?

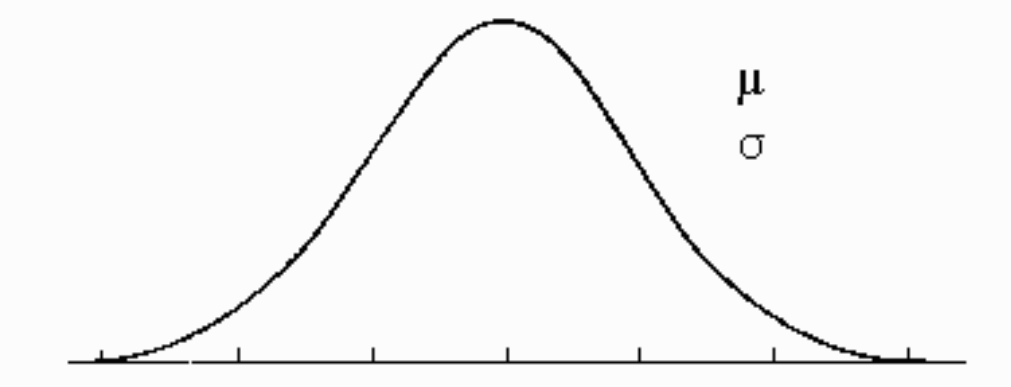
$Math\ N{\sim}(68, 4)$



$History\ N{\sim}(73, 7)$

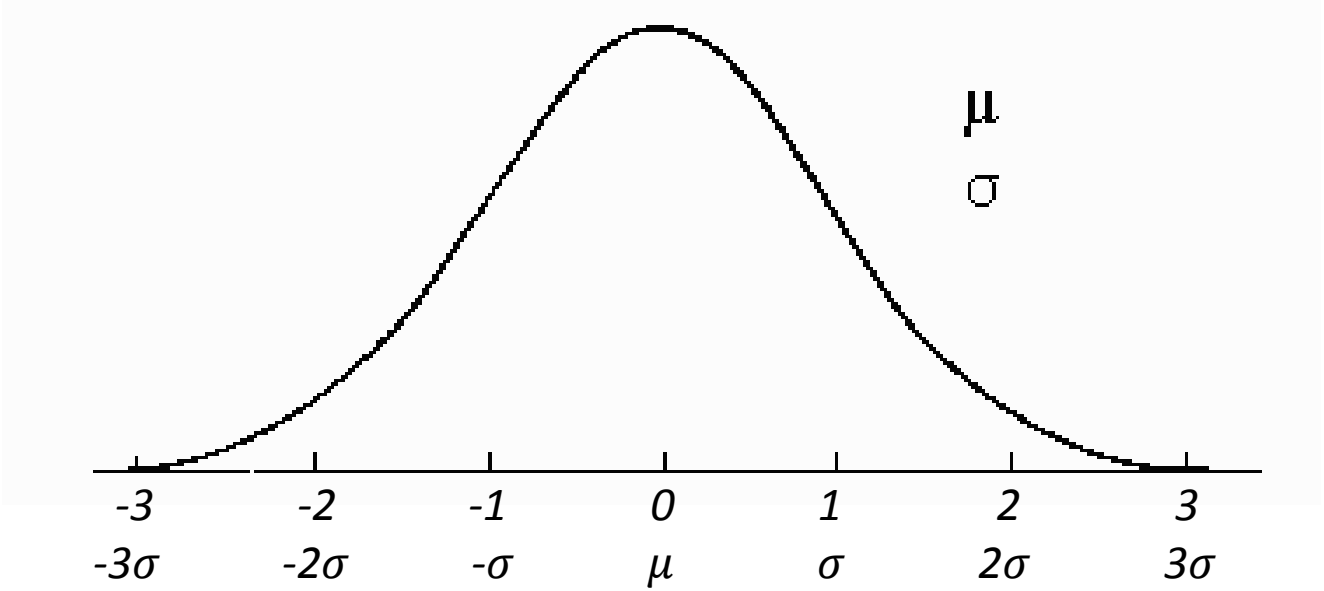


- Find the z score of each:



Standard Normal

- $X \sim N(\mu, \sigma)$
- $Z \sim N(0, 1)$
- Z-score $\frac{x-\bar{x}}{s}$ or $\frac{x-\mu}{\sigma}$

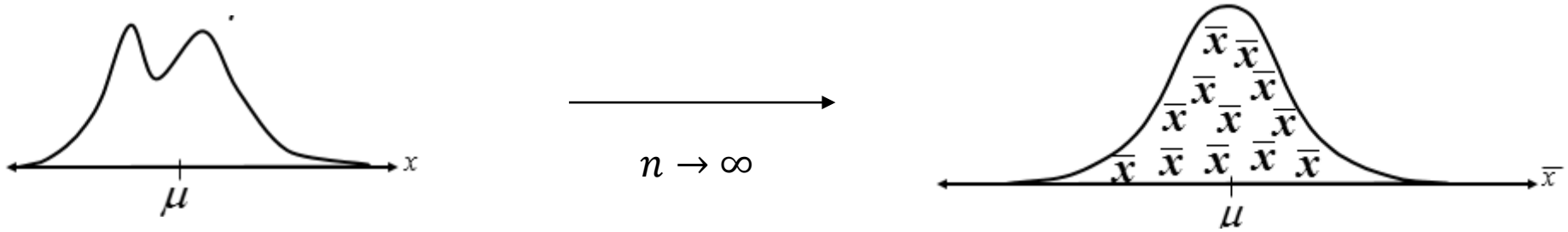


Find the standard normal score for a B, 80%, given that $N \sim (70, 10)$.

- Z-score $\frac{x-\bar{x}}{s}$ or $\frac{x-\mu}{\sigma}$

Central Limit Theorem (CLT)

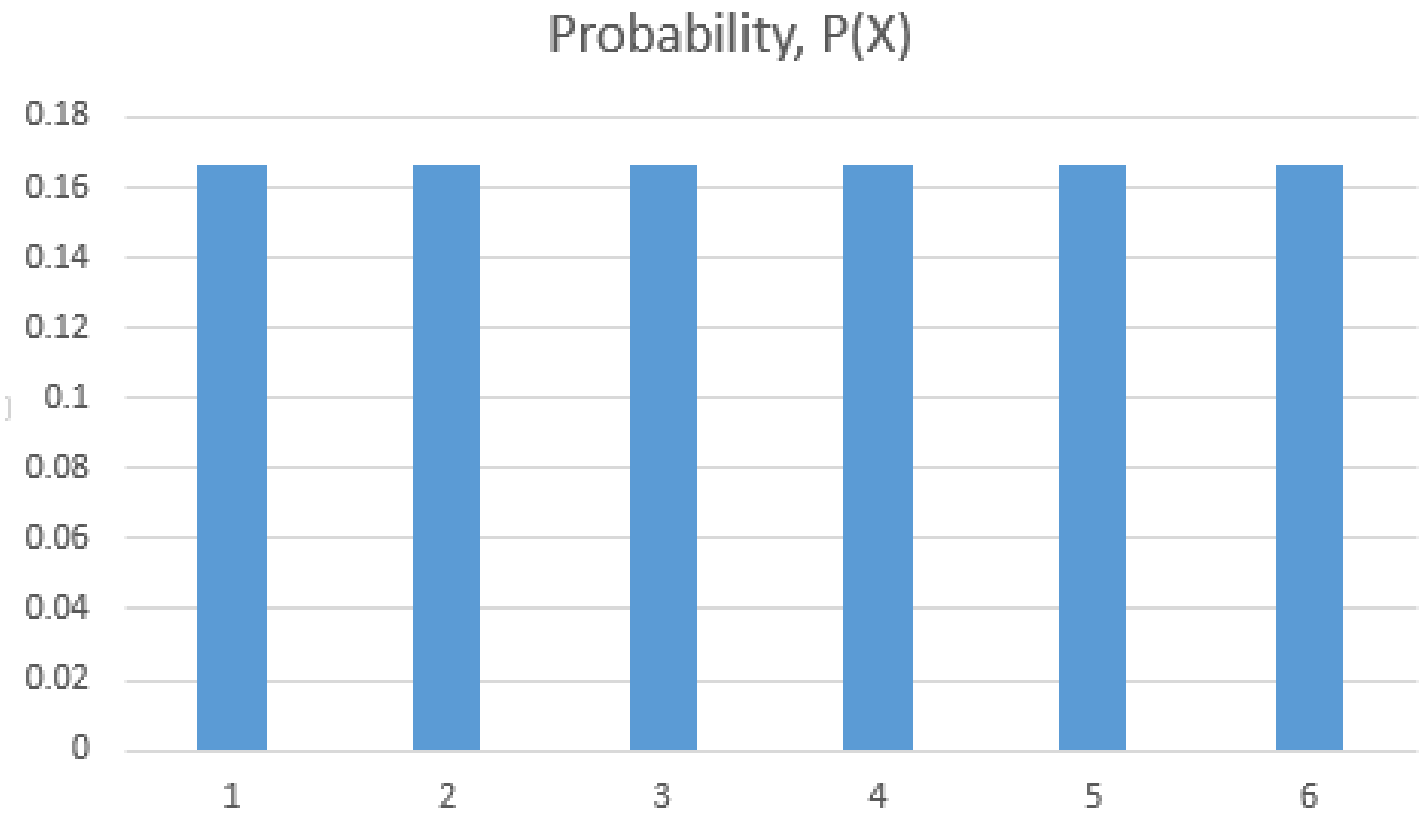
- Khan Academy CLT
- All sample distributions (shapes) approach a Normal Curve as $n \to \infty$



- *Sample means* have a Normal Distribution if n is large, generally n>30, even if the original distribution is not normal
- Sample is random and independent, equal chance of being selected
  - Uniform distribution

Rolling a Die example

- Random and Independent
  - Probability of each event is equal



Sample Mean Distribution, $n = 4$

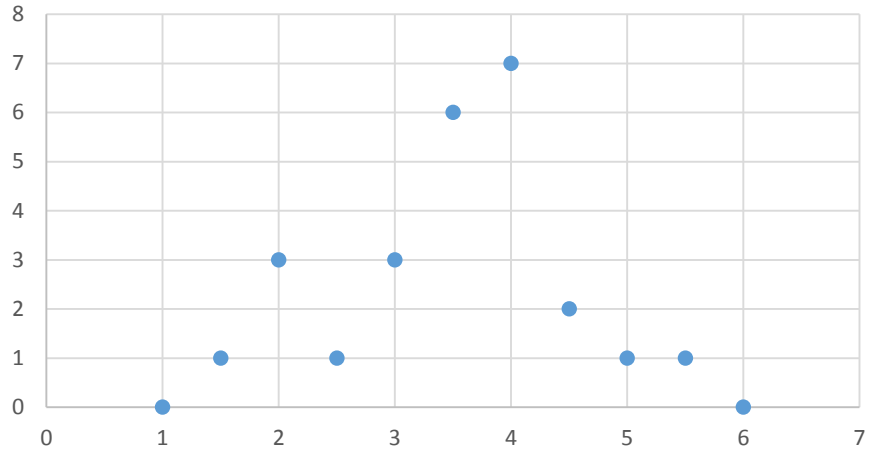
- 25 Samples, each sample has 4 trials

Sample Means, $\bar{x}'s$

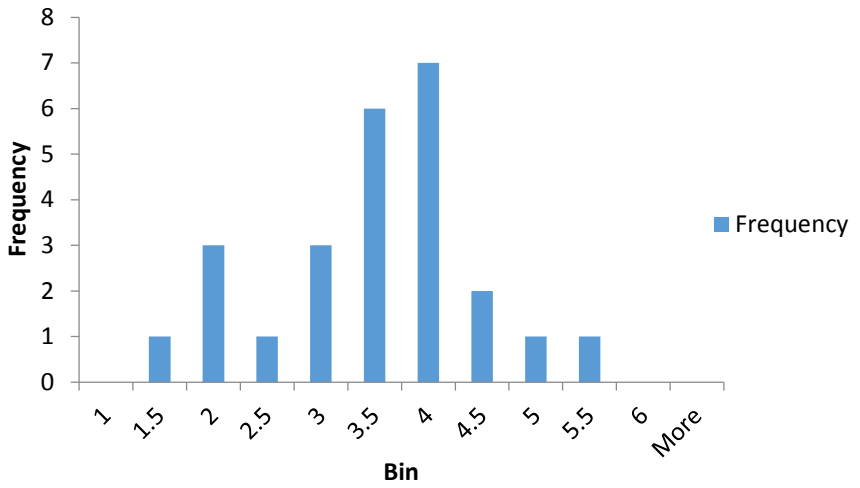
Sample	S1	S2	S3	S4	S5	S6	S7	S8	.....	S22	S23	S24	S25		
x1	4	6	1	2	4	6	5			4	4	1	5	2	
x2	4	6	3	2	6	5	6			1	3	3	6	3	
x3	4	1	4	6	4	5	1			3	4	3	6	4	
x4	3	5	3	3	2	3	3			4	5	1	4	4	
Average	3.75	4.5	2.75	3.25	4	4.75	3.75			3	4	2	5.25	3.25	3.833333
St. Dev.	0.5	2.380476143	1.258305739	1.892969	1.632993	1.089725	1.920286			745	0.707107	1	0.829156	0.829156	1.459078

- Plot the Sample Means, $\bar{x}'s$

Sample Means Dot Plot, n=4



Sample Means Distribution, n=4



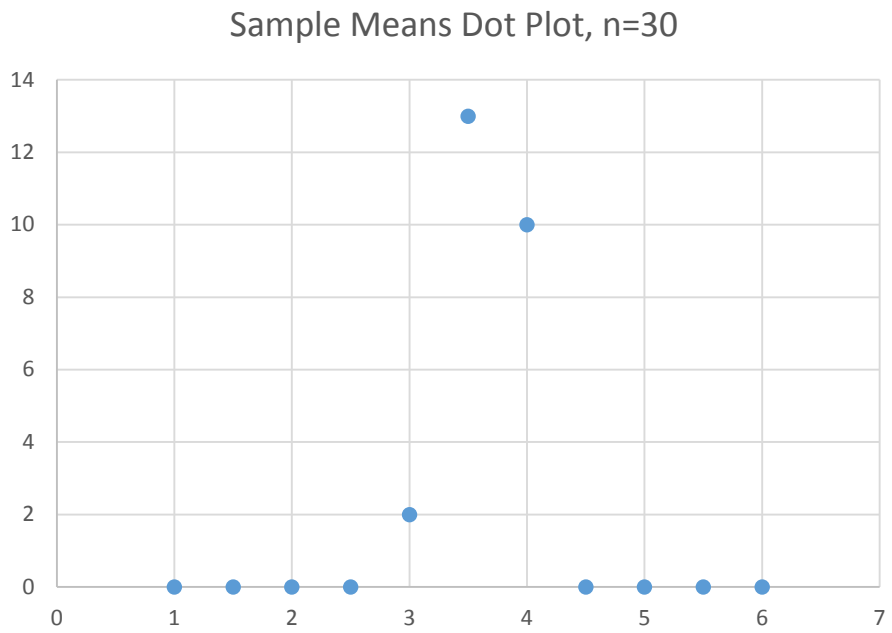
Sample Mean Distribution as n gets large, $n \geq 30$.

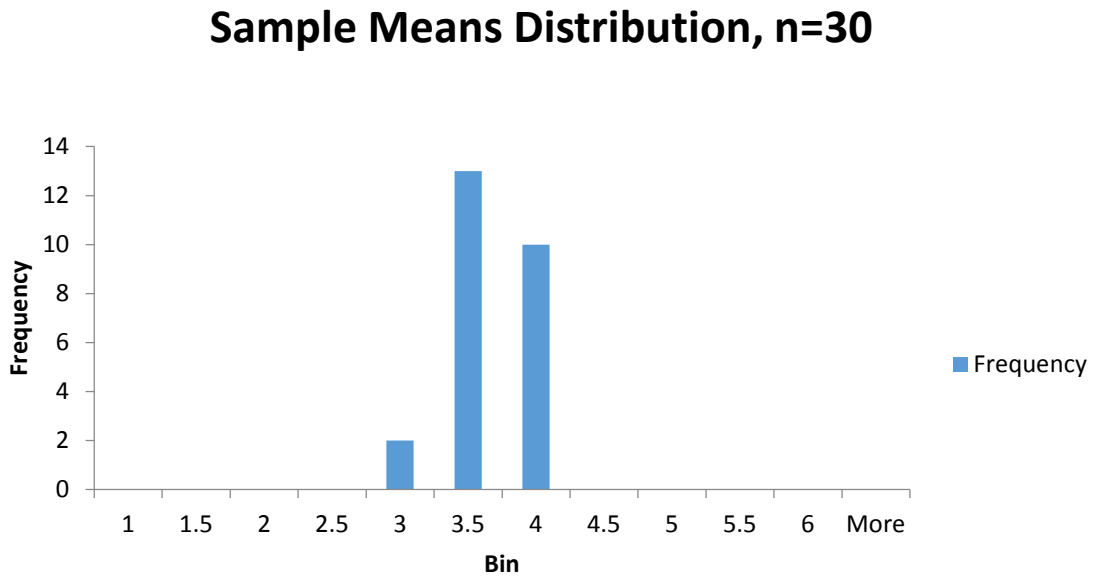
- 25 Samples, each sample has 30 trials

Sample Means, $\bar{x}'s$

Sample	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25
x1	4	4	4	4	1	2	1	1	5	3	1	5	6	5	6	2	2	2	5	5	4	3	2	4	6
x2	1	3	6	1	3	5	6	4	5	1	2	5	3	1	1	3	5	3	3	3	1	1	5	3	6
x3	5	4	2	2	1	3	1	1	4	3	3	4	3	2	5	3	2	3	5	1	2	4	4	1	4
x4	6	2	3	5	5	6	1	2	5	4	3	6	5	1	6	3	4	2	5	1	4	4	5	4	5
x5	2	5	1	3	2	1	4	1	6	6	1	6	5	2	2	2	3	1	1	4	2	1	4	2	3
x6	5	2	4	1	1	3	1	6	3	1	5	1	3	5	6	1	6	4	3	5	4	6	5	5	4
x7	3	4	2	1	2	5	5	3	5	1	6	3	3	5	4	5	3	5	1	6	4	6	5	4	4
x8	4	1	1	5	2	6	2	3	3	5	3	6	5	6	2	4	1	2	2	2	3	3	1	3	1
x9	4	6	3	5	6	2	1	4	2	6	4	3	6	6	5	6	2	2	6	1	5	1	4	3	5
x10	2	4	2	5	5	6	1	5	6	3	2	3	2	5	4	2	2	2	5	1	5	2	5	1	6
x11	5	4	3	5	2	3	6	4	5	3	5	3	3	5	6	3	6	3	5	4	4	5	1	2	5
x12	3	5	3	4	6	6	1	3	5	4	3	3	5	5	5	5	5	3	3	1	2	2	1	1	1
x13	2	5	4	1	4	6	5	4	1	3	3	6	2	2	5	5	2	3	4	1	2	5	4	5	5
x14	5	3	6	5	3	4	3	3	6	2	3	3	5	6	5	1	6	3	2	6	4	5	2	2	2
x15	2	6	5	6	5	3	4	5	6	2	3	5	5	4	5	1	5	3	1	5	5	3	5	2	2
x16	5	2	6	2	6	4	1	5	3	1	5	6	5	1	1	3	6	4	3	5	5	3	5	6	1
x17	1	3	5	4	5	3	2	4	2	4	1	6	4	3	6	3	2	6	3	2	5	5	3	4	4
x18	2	4	1	1	1	3	6	5	5	3	5	5	1	1	1	3	3	4	6	2	3	6	2	5	6
x19	4	6	6	5	1	5	5	1	5	3	2	6	6	2	4	2	2	2	5	3	6	2	3	5	3
x20	1	3	1	1	6	5	4	1	2	4	3	6	6	6	1	3	6	2	3	4	6	2	5	1	4
x21	5	4	6	4	3	3	4	3	6	1	3	4	3	5	2	4	1	5	6	3	3	3	5	5	4
x22	2	6	4	4	4	3	4	6	1	5	6	2	3	6	4	2	5	3	2	6	1	3	5	2	2
x23	3	5	4	2	6	4	3	1	4	3	1	1	2	3	3	4	5	2	1	4	5	5	1	4	1
x24	2	3	1	1	3	1	3	4	2	5	3	6	3	6	6	1	2	4	2	5	3	4	5	6	4
x25	3	6	4	4	1	6	3	5	3	5	6	3	6	5	2	3	3	4	3	3	6	6	4	2	6
x26	5	3	6	6	3	1	5	2	1	2	5	3	4	6	1	5	1	2	4	2	4	1	6	3	1
x27	3	4	5	1	4	2	5	5	1	1	5	2	1	2	5	3	1	6	1	2	1	3	2	2	3
x28	1	4	4	6	6	4	1	1	5	4	4	3	1	1	6	1	3	1	1	6	5	3	6	5	5
x29	5	6	1	3	6	1	4	2	6	5	6	6	4	5	6	5	3	3	2	6	5	4	5	2	4
x30	5	5	4	3	3	3	2	3	3	3	2	6	6	2	2	3	3	3	6	1	5	2	5	1	5
Averag	3.333333333	4.066666667	3.566666667	3.333333	3.633333	3.566667	3.133333	3.233333	3.933333	3.266667	3.466667	4.233333	3.866667	3.8	3.9	3.1	3.333333	3	3.3	3.333333	3.8	3.433333	3.833333	3.166667	3.73333	534667
ST Deve	1.538770968	1.387961376	1.774983854	1.787909	1.921243	1.788227	1.775227	1.698245	1.760255	1.574318	1.613164	1.675926	1.6244	1.810954	1.909091	1.446756	1.78271	1.606238	1.454952	1.517712	1.612095	1.59921	1.59921	1.700575	1.670559

Plot Sample Means, $n \geq 30$.

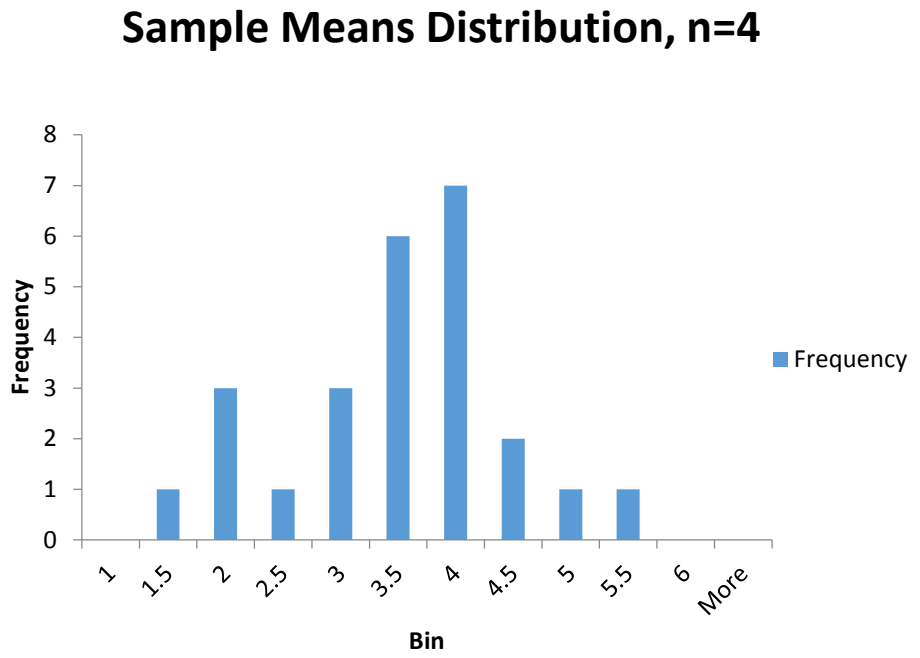


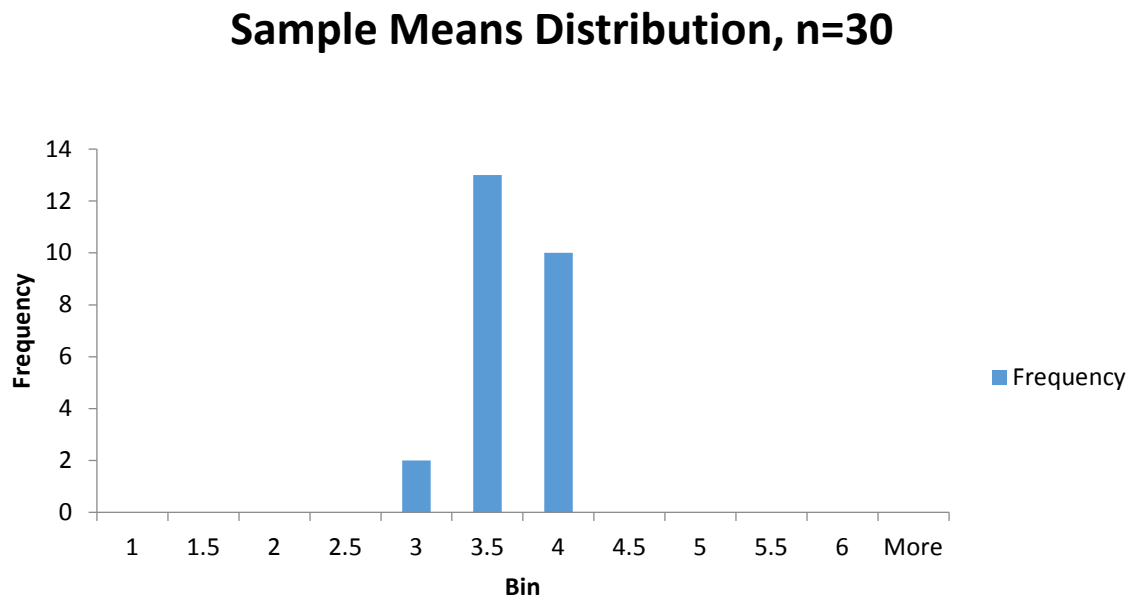


Compare Stats and Distributions of Sample Means as n → ∞…

CLT Roll Die Example

	Population	Sample, n=4	Sample, n=30
Average	3.5	3.833333333	3.534666667
Standard Deviation	1.707825128	1.459078205	1.670559199





What is Normal?

What is Normal with TInspire

Statistics on Geogebra

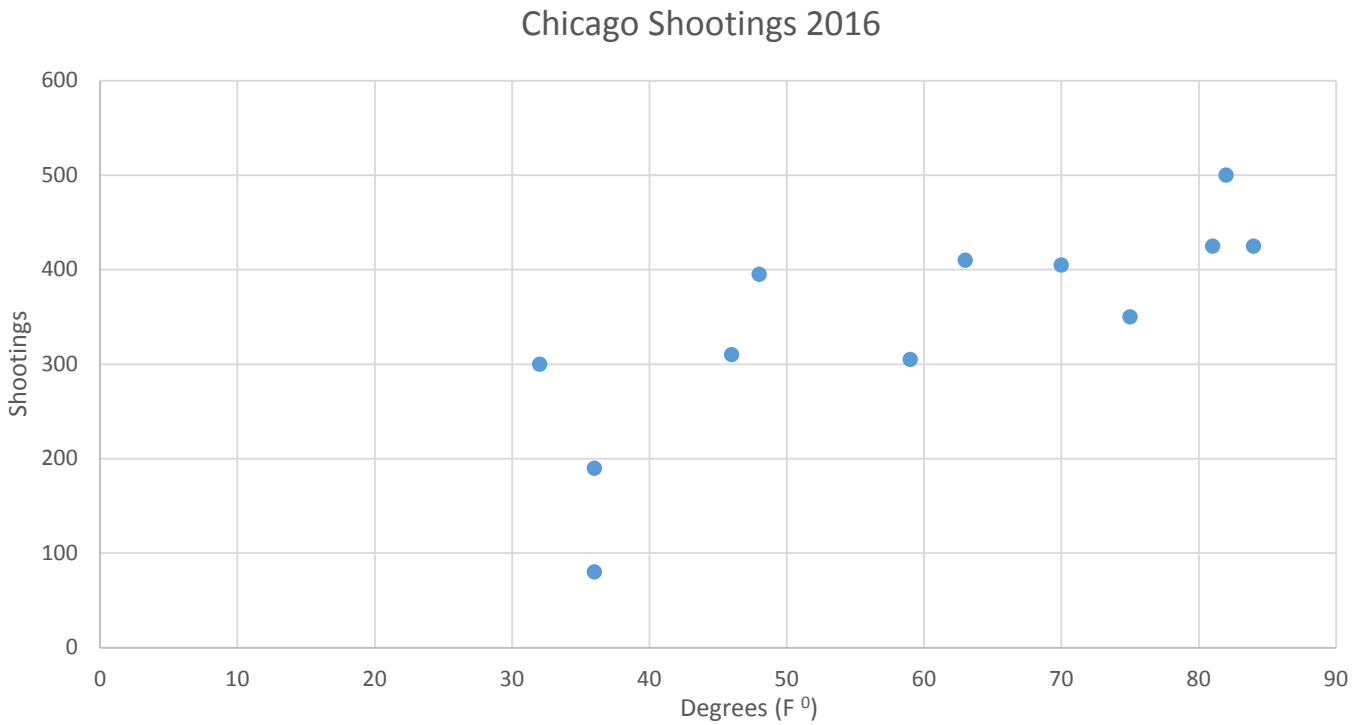
-

Big Picture, part 4: Linear Regression

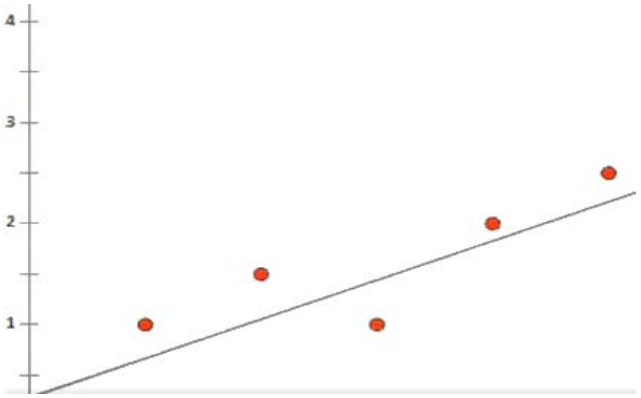
- Add example
- Explain stats by hand, intuitive approach
  - Correlation coefficient
  - Coefficient of determination
  - Least Squares

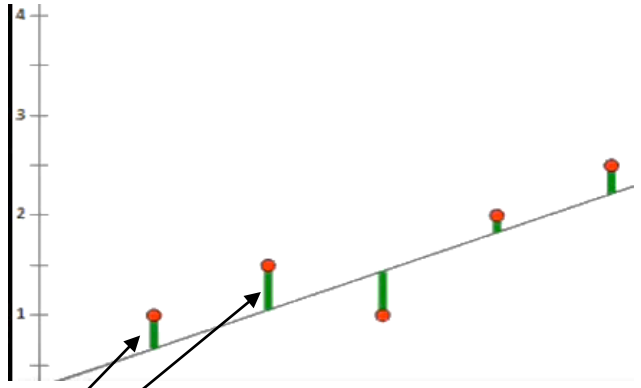
Are Temperature and Shooting Associated?

Month	Chicago Average High	Chicago Shootings
January	32	300
February	36	190
March	46	310
April	59	305
May	70	405
June	81	425
July	84	425
August	82	500
Septembe	75	350
October	63	410
Novembe	48	395
Decembe	36	80



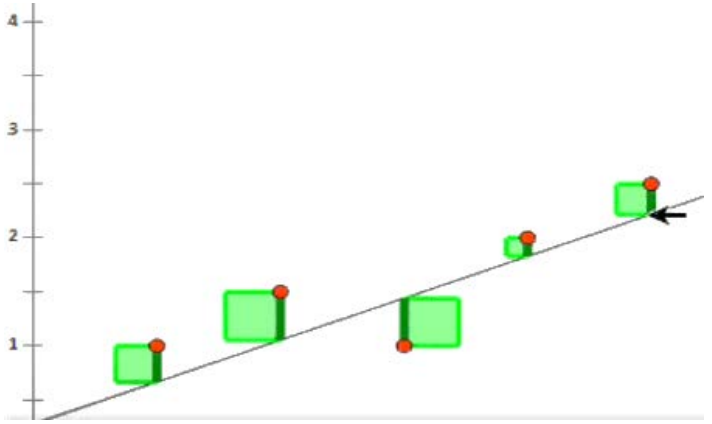
Find a Line of Best Fit: Least Squares Method

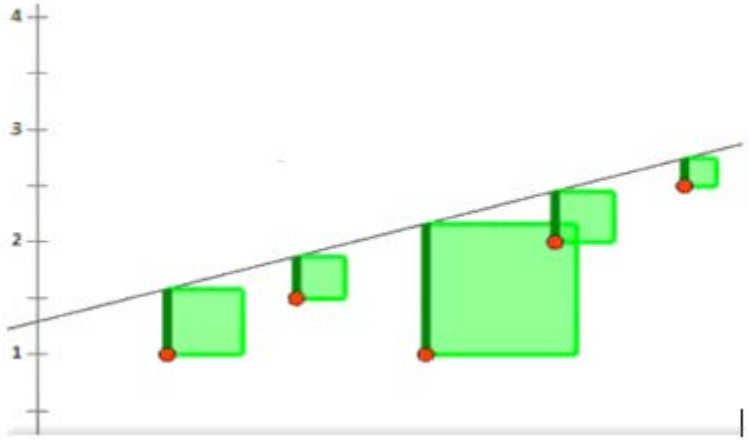




r, residuals

Which is the best model to fit the data?



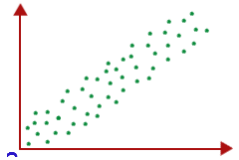


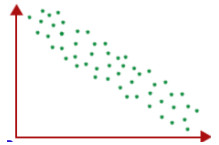
How well does the model fit the data?

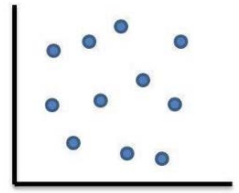
How much VARIATION?

Correlation Coefficient, r

- $[-1, 1]$
- Strength of (x, y) relationship
  - Strong if close to ± 1
    - $+1$
    - -1
  - Weak if close to 0





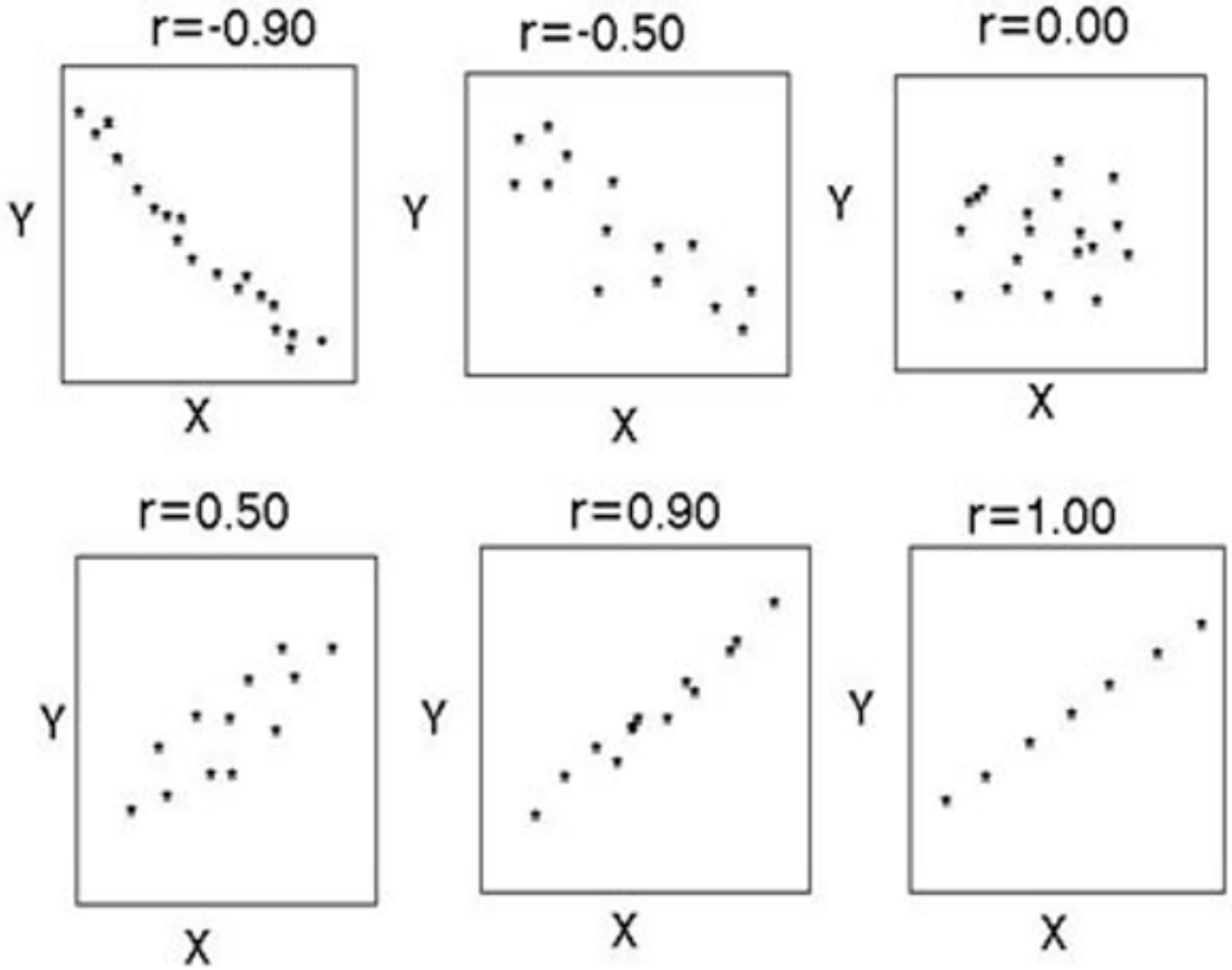


- $x \; and \; y$ have (strong/weak) (positive/negative) correlation.
- $x \; and \; y$ (are/are not) associated.

Coefficient of Determination, r^2

- $[0, 1]$
  - 0 means the model explains none of the variability around the mean
  - 1 means the model fits the data perfectly
  - Closer to 1 means more useful the model
- % of variation is y that can be explained by variation in x
  - % of the y can be predicted by x
  - If just $\bar{y}$ is used to predict y, then we are saying x does not contribute information about y

Approximate the Correlation Coefficient based on the graph and classify at Strong, Moderate, or Weak



Why not compute r^2 by hand?!

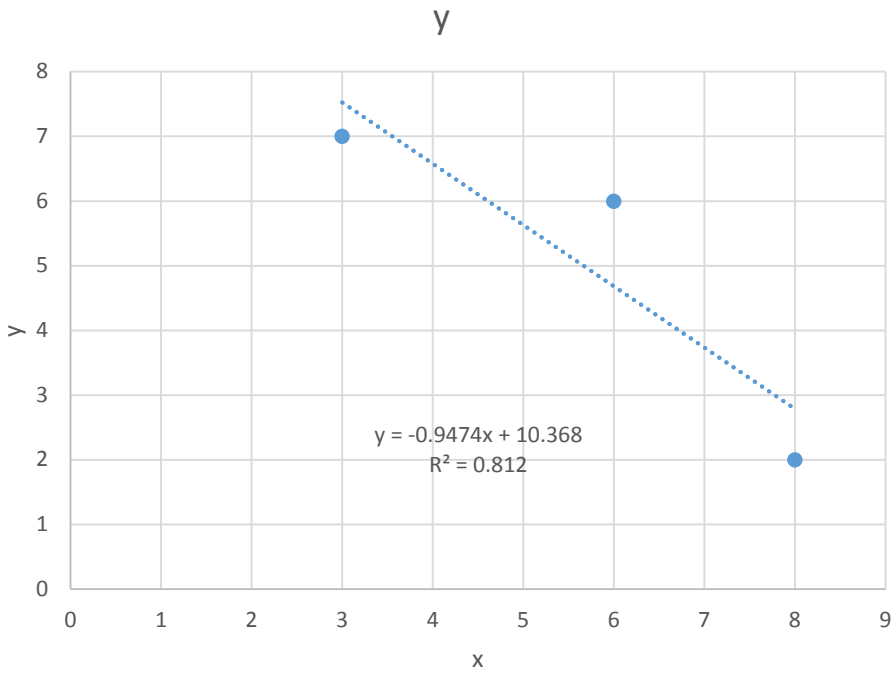
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

	x	y	(x-x bar)	(y-y bar)	(x-x bar)*(y-y bar)
	3	7	-2.66667	2	-5.333333333
	6	6	0.333333	1	0.333333333
	8	2	2.333333	-3	-7
Average	5.666667	5		Sum=	-12

	x	y	(x-x bar)	(y-y bar)	(x-x bar)*(y-y bar)
	3	7	-2.66667	2	-5.333333333
	6	6	0.333333	1	0.333333333
	8	2	2.333333	-3	-7
Average	5.666667	5		Sum=	-12
St Dev	2.516611	2.645751			
r	-0.90113				
r^2	0.81203				

$$r = \frac{-12}{(3-1)(2.52)(2.65)} = -.90$$

$$r^2 = .812$$



Question 4: Are Education Attainment and Salaries Associated? (Need to add Excel Tech notes)

Linear Reg Education

Linear Reg Ed Data

Favorite Tech/Apps

- Wolfram Mathematica
- Khan Academy
- Geogebra
- Desmos
- TI Inspire
- Excel
- Google Sheets

Data is the Answer to Everything!

- FALSE!
- But it could help if it is
  - Unbiased
  - Random
  - A correct model is selected to represent the data
  - Then data could help us search for the answers and truths.
- Even with best efforts by well respected institutions, random data is difficult to obtain.

Appendix B

Names _____

Representing Data

Materials: 20 pennies per group, Double sided tape, Poster Board

Question: How many years does a penny stay in circulation?

Part I. Collect and Represent the Data in at least two different formats.

Q1. Describe some of the pros and cons of the different ways to display data.

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Q2. What is your preferred way to display data? Explain.

Part II. After you collect your data, place your penny in the class Penny Poster in the front of the classroom to create a dot plot.

Part III. Analyze the Data, SOCS

SOCS	**Your Sample, n=20**	**Class Sample, n=**
Shape		
Outliers		
Center		
Spread		

Q3. State at least three things do you notice or wonder about the age of the pennies.

Q4. Compare and contrast your sample and the class sample.

Part IV. Conclusions

Q4. What are some conclusions you can draw about the age of the pennies?

Q5. What are the limitations of this study?

Q6. Can you generalize the above conclusions to the entire penny population? Justify your answer.

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Appendix B

National Council of Teachers of Mathematics, Principles to Actions Ensuring Mathematical Success For All, 2014, Steven Leinwand, Daniel J Brahier DeAnn Huinker RobertQ Berry III, Frederick L. Dillon, Matthew R Larson, Miriam A Leiva, W. Gary Martin, Margaret S Smith

Appendix C

Technology Notes for *TI nspire*

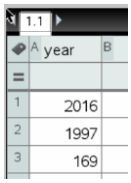
(How Many Years doe a Penny in Circulation?)

Question: How many years does a penny stay in circulation?

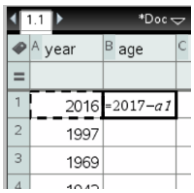
1. Compute Descriptive Statistics

Select 1. New Document, 2. List and Spreadsheets

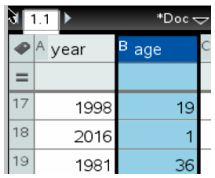
Enter the data Column A (Option: Name Column A "year" by selecting the top cell)



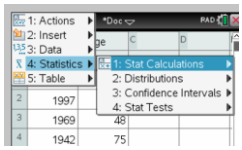
Enter in cell B1: $= 2017 - a1$ (Instead of typing a1, select the a1 cell), press Enter

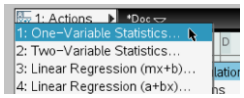


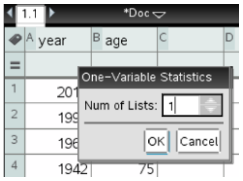
Select B1 to highlight**, hover arrow** over bottom right corner of cell until **plus sign appears, click, and drag down** until cell b20. (2017-a# will repeat). Click another part of the page to remove highlight from column B.



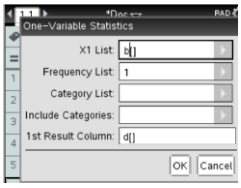
Select Menu, 4. Statistics, 1. Stat Calculations, 1. One Variable Statistics, OK



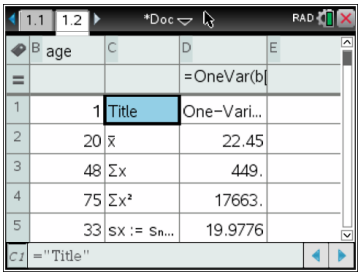




Enter in X1 List: $b[\,]$, **Select** OK.



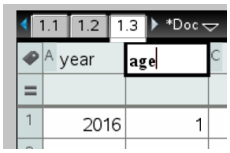
Results:



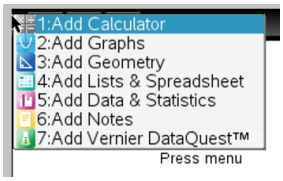
Note: For more advanced use, become familiar with the commands, $= OneVar(b[],1)$

For Results in more reader friendly format:

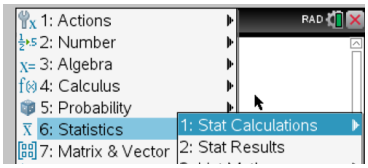
Select "age" cell, or cell at the top of column B, **Press** ctrl, sto→, and **type** age

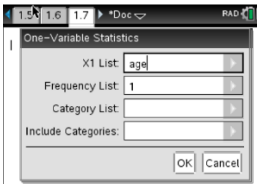


Press cntl, +page, 1: Add Calculator.

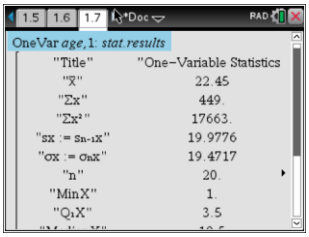


Press Menu, 6: Statistics, 1: Stat Calculations, 1: One variable Statistics, **Select** OK, **Enter in X1 List:** age **Select** OK,



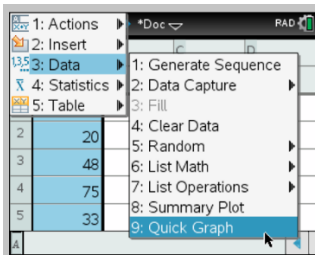


Results:

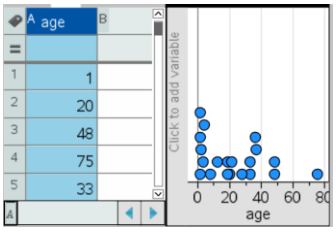


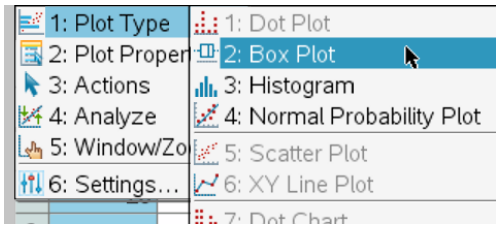
2. Represent the Data

Select and Highlight "age" column by clicking on the right line of the cell, **Select** Menu, 3: Data, 9: Quick Graph



Results: Change the graph type. **Click on graph, Select** Menu, 1: Plot Type, and then 2: Box Plot or 3: Histogram





To Undo Split Screen:

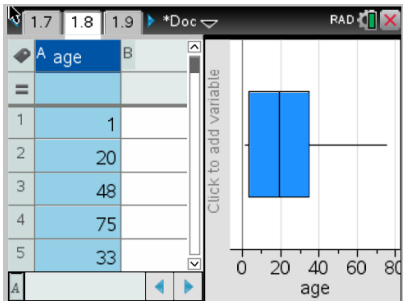
Doc,

5: Page Layout,

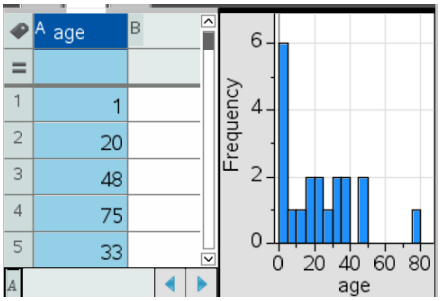
8: Ungroup

More Results in Split Screen:

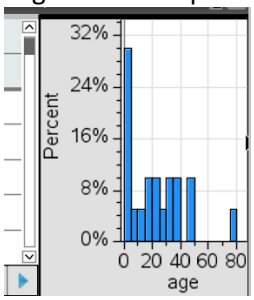
Box Plot



Histogram, Frequency Chart

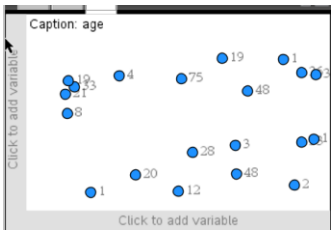


Relative Frequency Chart (change 4:Scale to percent)

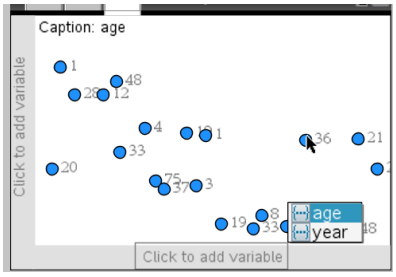


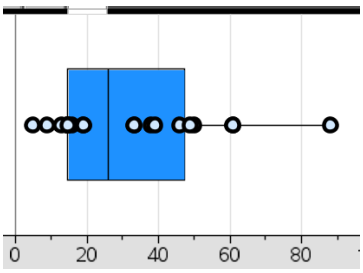
Results on a New Page:

Press ctrl, +page, 5: Data & Statistics, Enter.



Click to add variable, Select age, Enter. For more graph types, **Press** Menu, 1: Plot Type, 2: Box Plot or 3: Histogram.





Handy Tips:	
Ctrl Z	Undo
Ctrl Menu	Right Click
Ctrl Arrow	Move between tabs
Doc, 5: 5:	Delete a tab
Doc, 5: 8:	Undo Split Screen

Appendix D

Names _____

Sampling

Question: What is the readability of the Gettysburg Address?

Directions: Average word length is one of several measurements to help determine the readability of a passage. Find the average work length of the Gettysburg Address, a historic and world famous speech by Abraham Lincoln given at the Cemetery for Gettysburg in an attempt to give meaning to the events that took place one of which being over 50,000 causalities.

The Gettysburg Address

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate—we can not consecrate—we can not hallow—this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

Part I.

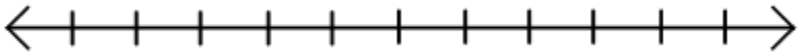
1. Select a random sample of 5 random words from The Gettysburg Address and circle them. Represent the data in a table (A) and dot plot (B).

Table (A)

n = 5	1	2	3	4	5
Word					
# of letters					

Appendix D

Box Whisker (B), n =5



(Based on the graph, estimate the mean.)

Indicate the following:

Observational units:

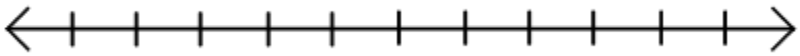
Variable: Continuous or Discreet

Type of Variable: categorical/numerical Sample

Your sample mean (average): $\bar{x} =$

2. At the front board, make a class dot plot (C), put your sample average on a dot plot with the rest of the class

Dot Plot (C)



(Based on the class sample dot plot, estimate the mean)

Now, compute the class sample mean (use *Google Sheets*), $\bar{x} =$

3. The population average (mean), μ is _____. Plot the population average on your Dot Plots (B) and (C). Note the population, N, of the Gettysburg Address is 268 words.

4. What are your observations about the sample means compared with the population mean?

   a. Were the sample means was accurate?

   b. Was one of the sample means more accurate than the other? If so, why do you think?

c. List some ways to improve your sampling so the sample mean(s) will be more accurate.

The **central limit theorem** states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough, $as\ n \to \infty$.

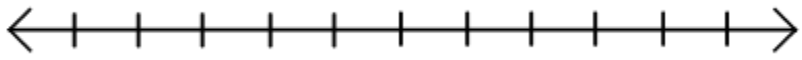
Part 2.

5. Use a computer to generate 5 random numbers between 1 and 268. It is ok to have a repeating number. Use The Gettysburg Address numbered word chart to find the corresponding word to the computer generated random number.

n = 5	1	2	3	4	5
Generated Random Number					
Word					
# of letters					

Your Sample Average (mean), $\bar{x}$:

6. Combine your sample average, $\bar{x}$, with the class on dot plot on the front board. (D).



7. By changing our sampling method change the results of our class sample mean? What conclusions can you make about sampling?

Sampling Frame

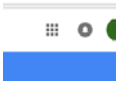
1	Four	55	We	109	cannot	163	for	217	they
2	score	56	are	110	dedicate	164	us	218	gave
3	and	57	met	111	we	165	the	219	the
4	seven	58	on	112	cannot	166	living	220	last
5	years	59	a	113	consecrate	167	rather	221	full
6	ago	60	great	114	we	168	to	222	measure
7	our	61	battlefield	115	cannot	169	be	223	of
8	fathers	62	of	116	hallow	170	dedicated	224	devotion
9	brought	63	that	117	this	171	hear	225	that
10	forth	64	war	118	ground	172	to	226	we
11	upon	65	We	119	The	173	the	227	here
12	this	66	have	120	brave	174	unfinished	228	highly
13	continent	67	come	121	men	175	work	229	resolve
14	a	68	to	122	living	176	which	230	that
15	new	69	dedicate	123	and	177	they	231	these
16	nation	70	a	124	dead	178	who	232	dead
17	conceived	71	portion	125	who	179	fought	233	shall
18	in	72	of	126	struggled	180	here	234	not
19	liberty	73	that	127	here	181	have	235	have
20	and	74	field	128	have	182	thus	236	died
21	dedicated	75	as	129	consecrated	183	far	237	in
22	to	76	a	130	it	184	so	238	vain
23	the	77	final	131	far	185	nobly	239	that
24	proposition	78	resting	132	above	186	advanced	240	this
25	that	79	place	133	our	187	It	241	nation
26	all	80	for	134	poor	188	is	242	under
27	men	81	those	135	power	189	rather	243	God
28	are	82	who	136	to	190	for	244	shall
29	created	83	here	137	add	191	us	245	have
30	equal	84	gave	138	or	192	to	246	a
31	Now	85	their	139	detract	193	be	247	new
32	we	86	lives	140	The	194	here	248	birth
33	are	87	that	141	world	195	dedicated	249	of
34	engaged	88	that	142	will	196	to	250	freedom
35	in	89	nation	143	little	197	the	251	and
36	a	90	might	144	note	198	great	252	that
37	great	91	live	145	nor	199	task	253	government
38	civil	92	It	146	long	200	remaining	254	of
39	war	93	is	147	remember	201	before	255	the
40	testing	94	altogether	148	what	202	us	256	people
41	whether	95	fitting	149	we	203	that	257	by
42	that	96	and	150	say	204	from	258	the
43	nation	97	proper	151	here	205	these	259	people
44	or	98	that	152	but	206	honored	260	for
45	any	99	we	153	it	207	dead	261	the
46	nation	100	should	154	can	208	we	262	people
47	so	101	do	155	never	209	take	263	shall
48	conceived	102	this	156	forget	210	increased	264	not
49	and	103	But	157	what	211	devotion	265	perish
50	so	104	in	158	they	212	to	266	from
51	dedicated	105	a	159	did	213	that	267	the
52	can	106	larger	160	here	214	cause	268	earth
53	long	107	sense	161	It	215	for		
54	endure	108	we	162	is	216	which		

Appendix E

Technology Notes for Google Sheets

(What is the readability of the Gettysburg Address?)

Open Google Sheets

1. Sign into gmail or create a google account.
2. Select Square icon in top right corner. 
3. Select Sheets Sheets . If not displayed, select "More" at the bottom, select "Even more form Google", then select Sheets.
4. Select Blank

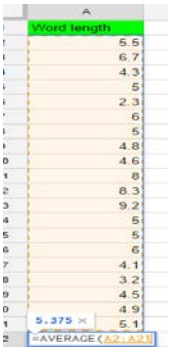
.

Find the mean, $\bar{x}$, of the class sample, n.

1. Enter data in column A…Heading and highlight are optional

	A
1	Word length
2	5.5
3	6.7
4	4.3
5	5
6	2.3
7	6
8	5
9	4.8
10	4.6
11	8

2. In an empty cell (under data), type " = Average(A2:A20)" or " = Average($highlight\ data$) ", press enter

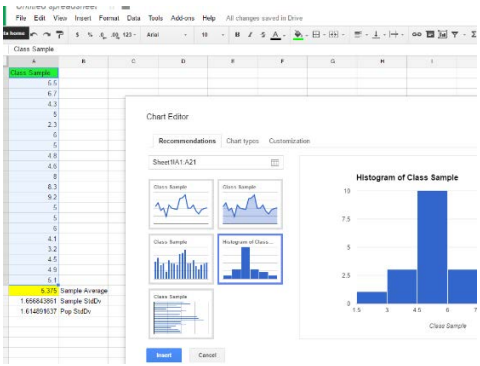


Google Sheets notes:

Sample standard deviation = Stddev($highlight\ data$)

Population standard deviation = Stddevp ($highlight\ data$)

3. Highlight data, Select Insert, then Chart, then Histogram, click on Insert

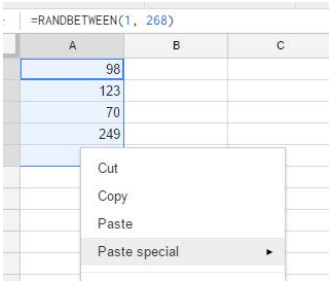


Explore Chart Features

Click on any part of the chart to change the title, font, color and many more settings.

Generate Random Numbers

1. Optional: Open a new sheet within your existing worksheet by selecting the plus sign, +, at the bottom left corner of the page.
2. In cell A1, type "= Randbetween(1, 268)", press Enter. Select cell A1, Hover arrow on bottom right corner of cell until a + appears, and Drag down until you have 5 random numbers.
3. Highlight data, Right Click to Copy, Select an empty cell, then Paste Special, select Paste Values Only. (This will stop from random numbers from recalculating.



4. Enter the means of the random samples from the class.
    a. Then repeat steps from previous section to mean, standard deviation and make a histogram.

Appendix F

Name _____

Who Is the Most Elite Athlete?

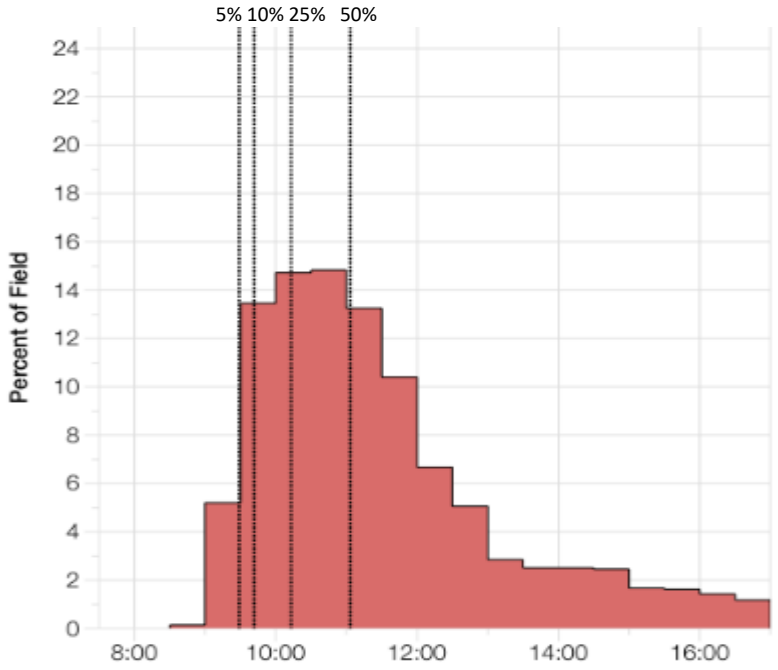
Q. Who is a more elite athlete, a cross country runner, a swimmer, or bicyclist?

1. What are some ways this question can me answered?

History of the Ironman

The athletes at the 1977 O'ahu Perimeter Relay designed the Ironman Triathlon to determine the most elite athlete. It included three events, a 2.4-mile swim, 112-mile bike ride, and 26.2 mile run. Let us begin analyzing data from these events.

2. Make some observations about the probability density functions, pdf, for the 2015 World Championship Ironman in Hawaii.



Shape

Outliers

Center

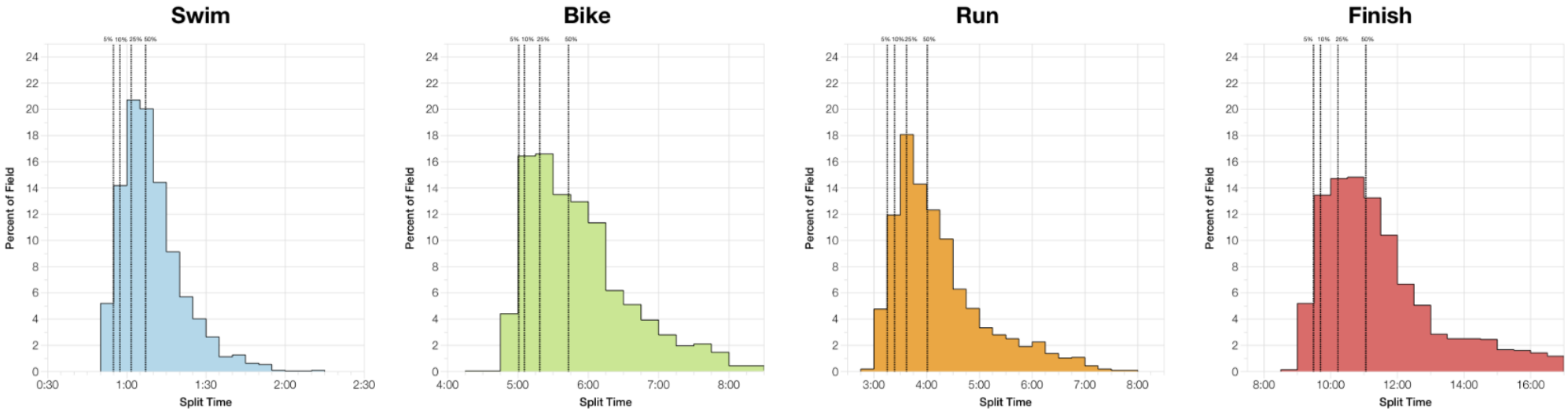
Spread

Appendix F

Note: Relative frequency chart

2016 Ironman World Championship

3. Decide which of the following a greater athletic triumph is, support your answer.



A. Finishing the swim portion of the race in 1:30 or the bike part in 5:00?

B. Finishing the bike part in 5 hours or finishing the run in 5 hours?

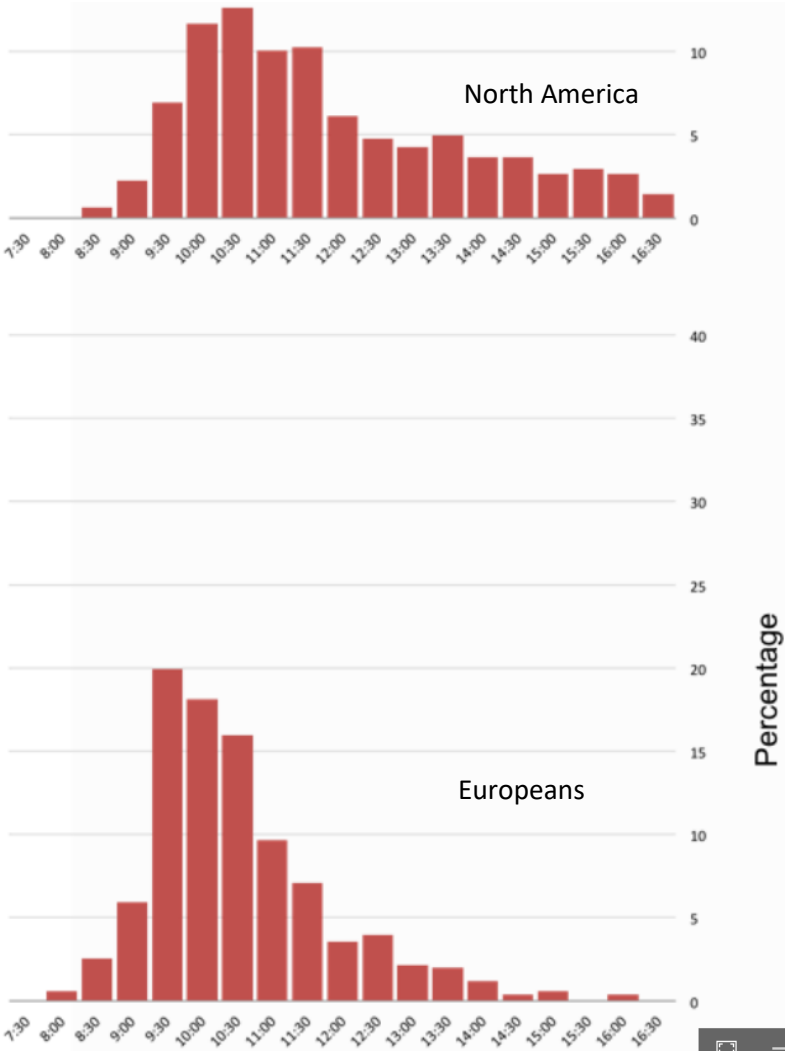
C. Finishing the swim in 2 hours or finishing the whole race in 12 hours?

4. If a triathlete completed the swim in 1:30, the bike in 6:00, in how much time would he have to complete the run to finish in the top 25%?

If a randomly chosen runner was selected, what is the likelihood they would finish in that time?

Appendix F

5. Are Americans or Europeans better athletes?



[Ironman Stats]

Appendix G

Name _____

What is Normal?

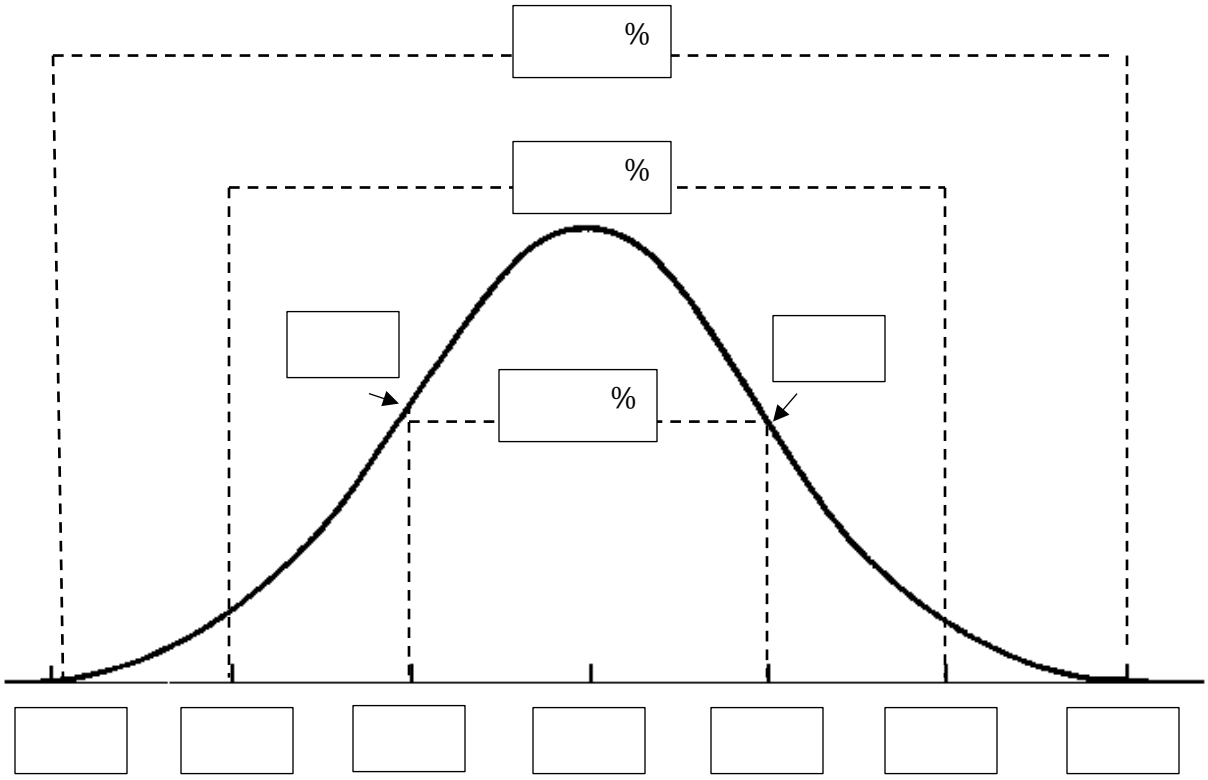
Guided Notes: Everything You Need to Know About Being Normal

Graphing Perspective. Label the key parts of the Normal Curve: $\bar{x}$

$\pm 3\ \sigma's$

$2\ points\ of\ inflection *$

$Empirical\ Rule$: $68\%, 95\%, 99.7\%$



Notation: ~(,)

Properties

-
-
-
-

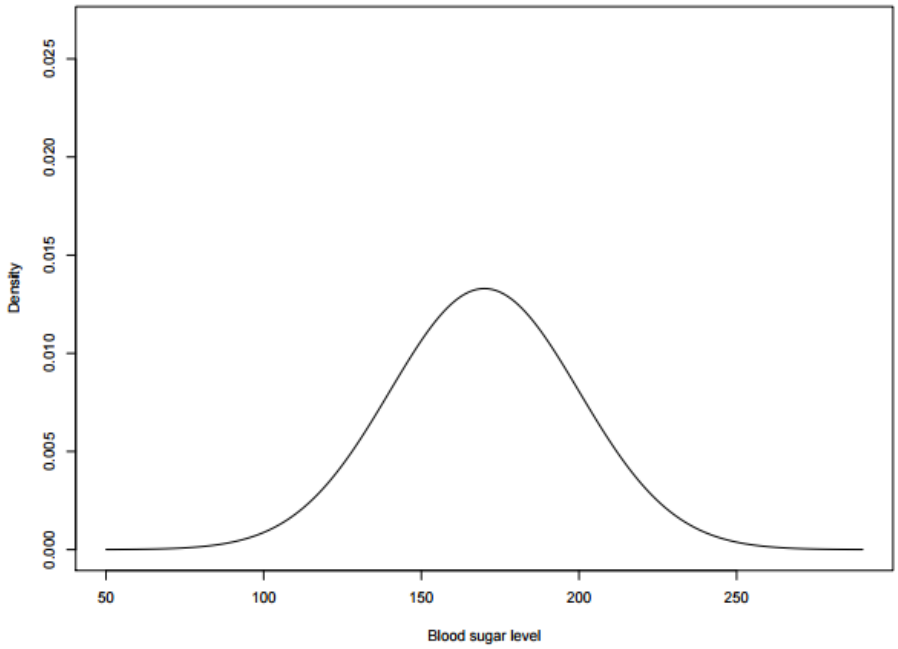
➢

Central Limit Theorem (CLT)

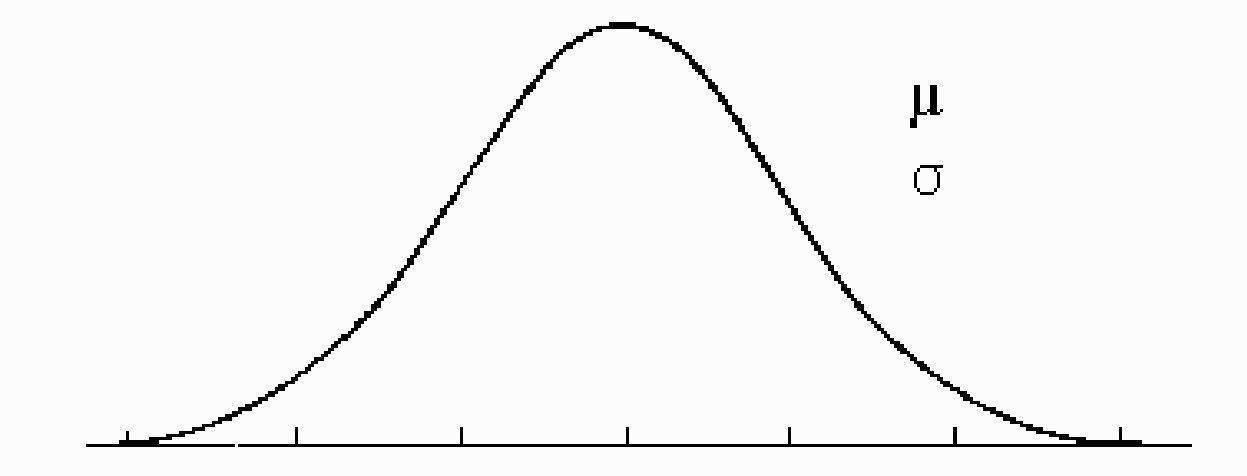
State in your own words and draw a picture to show what you mean.

Examples

1. Approximate the mean and standard deviation from the graph, $N \sim ($, $)$

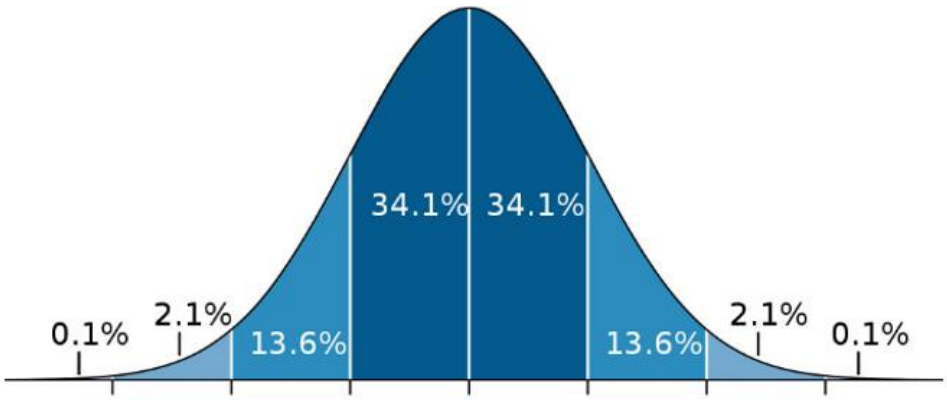


2. Label mean and standard deviations of the average highway speeds, given, $N \sim (\ 60 \ , 5 \ )$.



3. Men's shoe sizes are normally distributed given $N \sim(\ 10\ ,.5\ )$.

a) Label the graph with the mean and standard deviation.

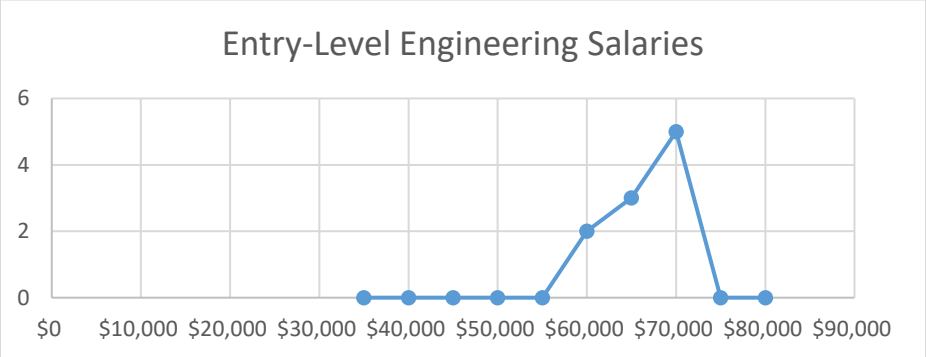


b) What percent of men have a shoe size larger than size 11?

c) If there were 45 men in this room, estimate how many would have a shoe size greater than 11.

d) If a man were randomly selected, what is the probability his shoe size is between size 9 and 10, $P(9 < X < 10) =$?

4. The average entry-level engineering salary is approximately $N \sim(\ \$63{,}000,\ \$4500)$.



a) What is the probability that a starting salary would be greater than \$70,000 $P(X \geq \$70{,}000)$?

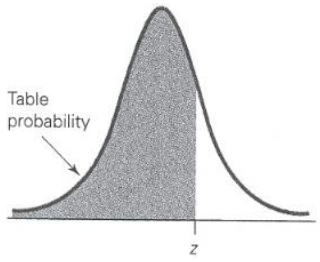
- Shade in the area on the graph.

- Compute the Z score, $\frac{x-\bar{x}}{s} = \frac{\quad -}{\quad\quad\quad\quad}$

- Use the Normal Distribution Table to find the $p - value$.

Appendix G

TABLE A.1 ■ Standard Normal Probabilities (for $z > 0$)



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706

Note: $(1 - p - value)$ *will give the area to the* ***right*** *of the z value.*

- Now use the TInspire

Steps

1. Add Calculator, Press Menu, 6. Statistics, 5. Distributions, 1. Normal CDF

Lower Bound: 70,000
Upper Bound: Very big number such as 1,000,000
μ: 63,000
σ: 4,500

- Use the TInspire to look at the area on the graph

Steps

4. Add Lists and Spreadsheets, Select the shaded cell "A=", Press Menu, 4. Statistics, 2. Distributions 2. Normal Cdf

Lower Bound: 70,000
Upper Bound: Very big number such as 1,000,000
μ: 63,000
σ: 4,500
Check Mark Draw Box

Select OK

Note: For better viewing: Select Doc, 5. Page Layout, 8. Ungroup

- Use the TInspire to answer what do the top 1% of entry-level engineers earn?

Steps

1. Add Calculator, Press Menu, 6. Statistics, 5. Distributions, 3. Inverse Normal

Area: 99%, .99
μ: 63,000
σ: 4,500

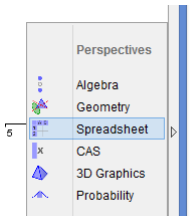
Introductory Statistics on Geogebra

This project will explore Introductory Statistics topics using tools in Geogebra to analyze data.

To get started…

1. Search Geogebra online and download.
2. Create an account (optional)
3. Open Geogebra if not automatically opened

4. Select on the right side of screen. Select Spreadsheet to view.



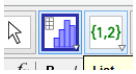
Please note: More views are available by selecting View at the top of the screen, such as Algebra, CAS, etc.

5. Enter Data in spreadsheet. Dataset is from US Department of Education. Please note: You can also highlight, copy, and paste any dataset online from most formats, i.e. Excel, Google Sheets, or lists.

Year	UW Sex Offense	IL Sex Offenses	Duke S Offenses
2001	14	14	8
2002	14	9	10
2003	13	12	6
2004	14	10	10
2005	14	8	10
2006	8	6	7
2007	7	9	5
2008	5	7	5
2009	4	5	13
2010	9	7	12
2011	24	5	7
2012	15	10	12
2013	29	12	6
2014	26	11	14

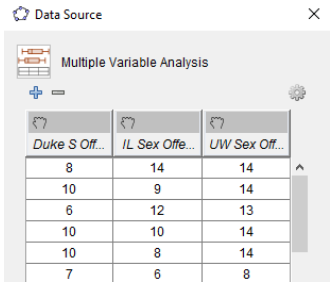
Appendix H

Multivariate Analysis

Highlight columns B, C, and D and Select Histogram icon 

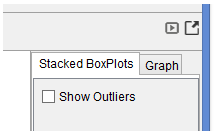
Select Multivariable Analysis.

In Data Source dialogue box select and check “Use Header as Title”



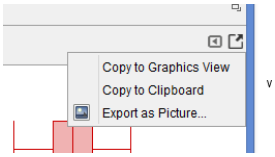
Duke S Off...	IL Sex Offe...	UW Sex Off...
8	14	14
10	9	14
6	12	13
10	10	14
10	8	14
7	6	8

Select Analyze.

Select the arrow in the Data Analysis Box. Note that Show Outliers is optional. 

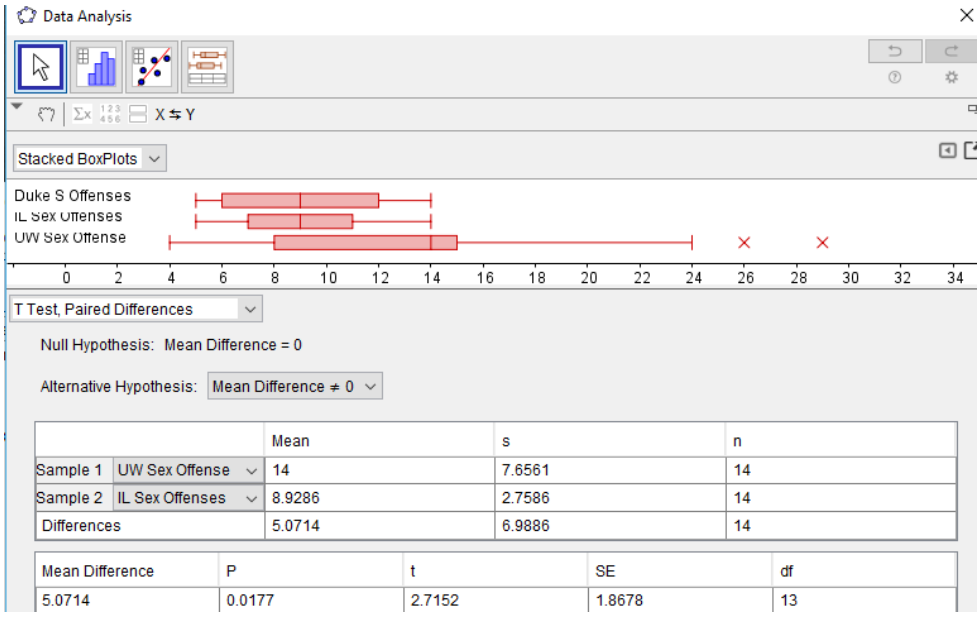
Select $\sum x$ to display statistics.

Select export arrow , and Select “Copy to Graphics View”



Change Color of Box Plots. Click on any one of the Box Plots, Select Object Properties, and Select Color.

Optional: In Data Analysis dialog box, Select Statistics to do a variety of tests, such as T tests, ANOVA, and compare any university from the



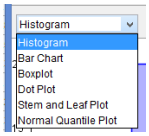
		Mean	s	n
Sample 1	UW Sex Offense	14	7.6561	14
Sample 2	IL Sex Offenses	8.9286	2.7586	14
Differences		5.0714	6.9886	14

Mean Difference	P	t	SE	df
5.0714	0.0177	2.7152	1.8678	13

Appendix H

One Variable Analysis

Highlight one of the columns of data, select the Histogram icon, and Select One Variable Analysis.

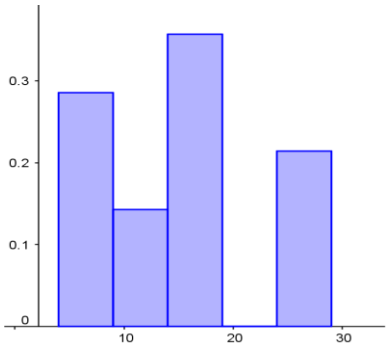
In the Data Analysis dialog box, click the down arrow button next to Histogram 

and explore some of the different graphics such as dot plots, stem leaf, etc. After looking at few different graphics go back to Histogram.

Drag the slider to change the number of classes, or width of the bins. Select the arrow,

Note the Classes and data on the class boundary will be put in the left bin. Note the Cumulative option.

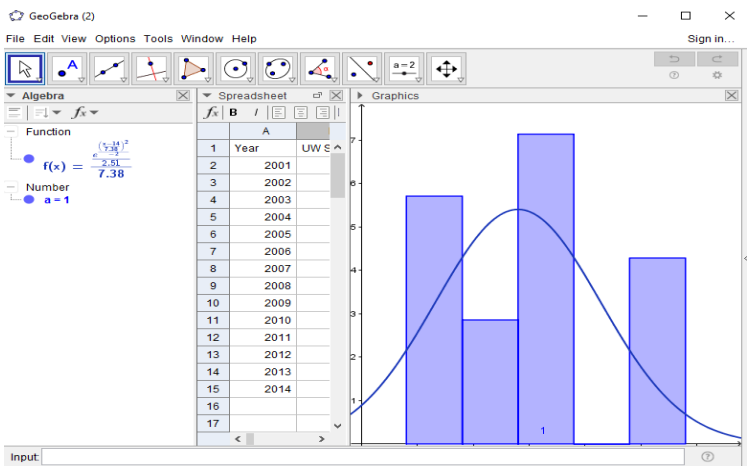
Check mark Relative Frequency and then Select Export Arrow .



Use the graph to approximate the percent of years were more than 20 offenses were reported. This could graph could also approximate the probability that a randomly selected year would have more than 20 offenses reported, $P(x \geq 20) =$

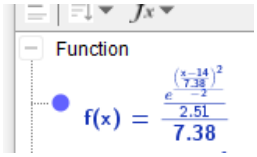
Let's say our data is continuous and we fit a normal curve to the data….Go back to the Data Analysis window. Check mark Normalized and Normal Curve and export .

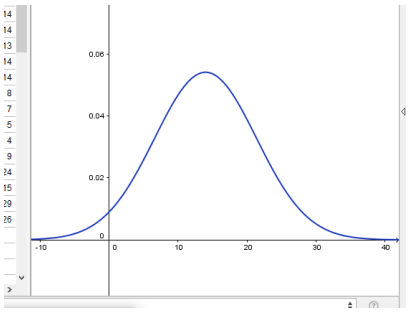
From the top of the main screen, Select View, Select Algebra. Again, Select View, Select Input Bar.



Appendix H

In the Algebra window, Click on the letter representing the histogram to hide the histogram. Note: f(x) is the equation of the curve.

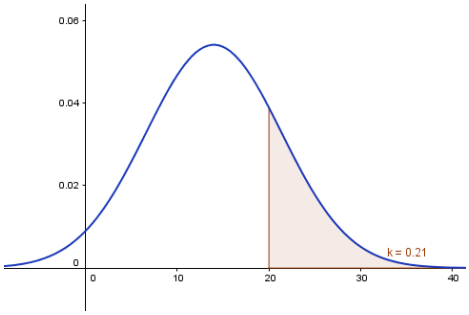




Example: Based on the US Department of Education's data set regarding sex offences on campus, what is the probability that more than 25 sex offenses will be reported, $P(x > 20)$ by

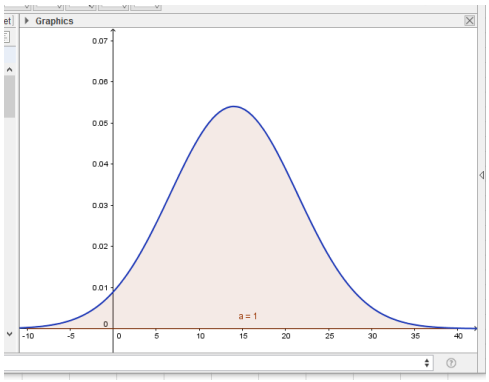
$$\int_{20}^{\infty} : \frac{e^{\frac{\left(\frac{x-14}{7.38}\right)^2}{-2}}}{\frac{2.51}{7.38}}$$

Calculate by entering the following in the Input bar: Intergral[f(x), 20, 100]



Integrate the whole function to demonstrate the area is 1 by entering:

Input: Integral[f(x), small number, large number] Try Intergral[f(x), -10, 50]. Note: Selecting too extreme values may stall the software.

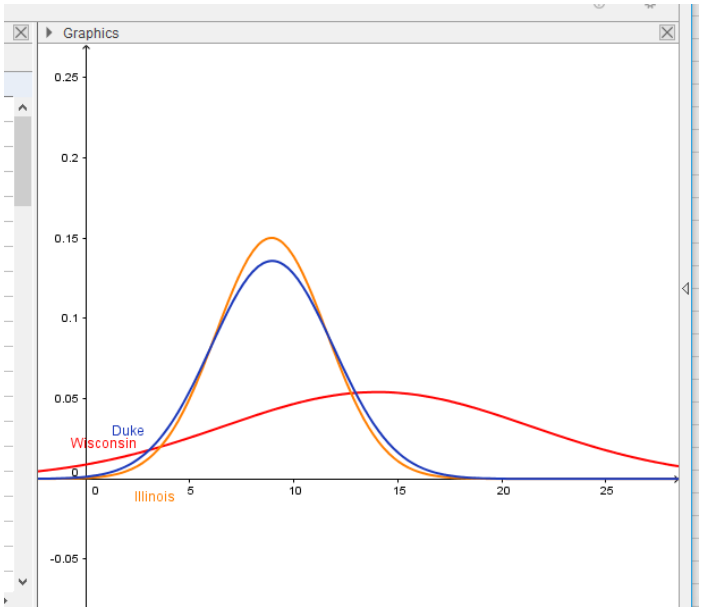


Appendix H

Explore the Graphics: Double Click on Curve, Select Object Properties, Select Basic, in the Caption Bar: Wisonsin. Select Color and change to Red.

Compare distribution from Duke and Illinois.

Go back to Spreadsheet, Highlight Column B for Illinois. Again Select Histogram and One Variable Stats. Setting for a Normalized Curve are probably still set, but if not select the arrow and again Check Mark Normalize and Normal Curve. (Uncheck Histogram since we will just compare curves.) Select export arrow , Select Copy to Graphics View. Again, Double Click on curve, Select Object Properties, Select Basic, in the Caption Bar: Illinois. Select Color and change to Orange. Repeat for Duke.



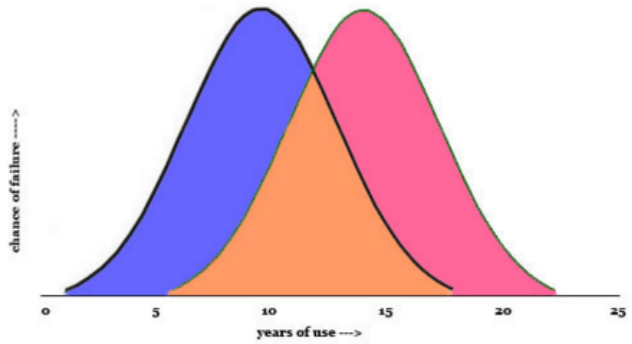
Make a note about the center and spreads for each school. What are your thoughts about the data? What can you conclude?

Optional Enrichment

Calculate $\int_{?}^{?} f(x) - g(x)$ Input: IntegralBetween[f(x), g(x), Lower Bound, Upper Bound]

Approximating Overlapping Coefficient (OVL)

Example: Consumer Application of an OVL for better understanding



Notice the three colored spaces? They each tell a different story, and their relative sizes are really important.

The blue area shows how likely it is that the cheap machine will fail before the expensive machine does.

The pink area shows something very similar: how likely it is that the expensive machine will last longer than the cheap machine.

To put it simply, you can combine the blue and pink areas – combined, they show the likelihood that the more expensive and reliable machine will outlast the cheaper and less reliable machine. There's about a 70% chance of that.

Of course, there's that orange area. That represents the chance that the cheap machine will actually last longer than the expensive machine. In this example, there's roughly a 30% chance of that happening.

Understanding the Reliability Bell CUrve

Paying more for reliability means that the orange area is your *worst* case scenario; buying the cheap machine means that the orange area is your *best* case scenario.

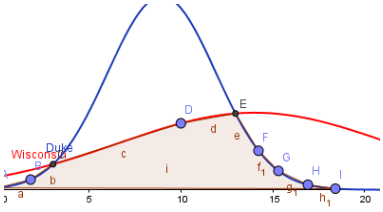
Appendix H

Example: Use two ways to find the area under the Overlapping Part of the Curves. One way will be to fit a polygon to the area and the second way will be to take the integral for the difference between the 2 curves.

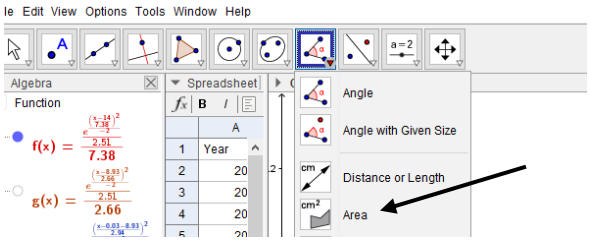
Optional: Get new data, find a dataset online that interests you from 2 different populations.

Enter the datasets in the spreadsheet and run the One Variable analysis on each. Export and Cope to Graphics View normalized curves.

To approximate overlapping area, select Polygon and try to closes match the vertices to the curve.



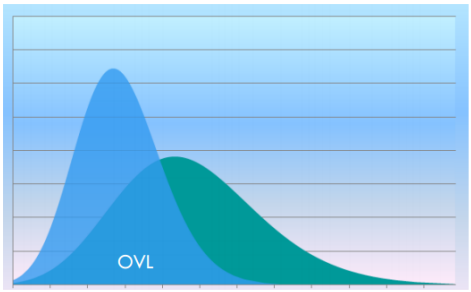
Select Area Tool.



What is the approximate area? (Should be between 0 and 1)

Now let's use the formula to see how closely our polygon matches the area. Here's the math behind calculating the OVL…

$$OVL = \int_{R_n} \min[f_1(\mathbf{x}), f_2(\mathbf{x})]d\mathbf{x}$$

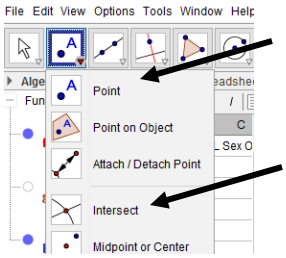


$$OVL = 1 - \left|\Phi(x_2) - F_2(x_2)\right| - \left|\Phi(x_1) - F_2(x_1)\right|$$

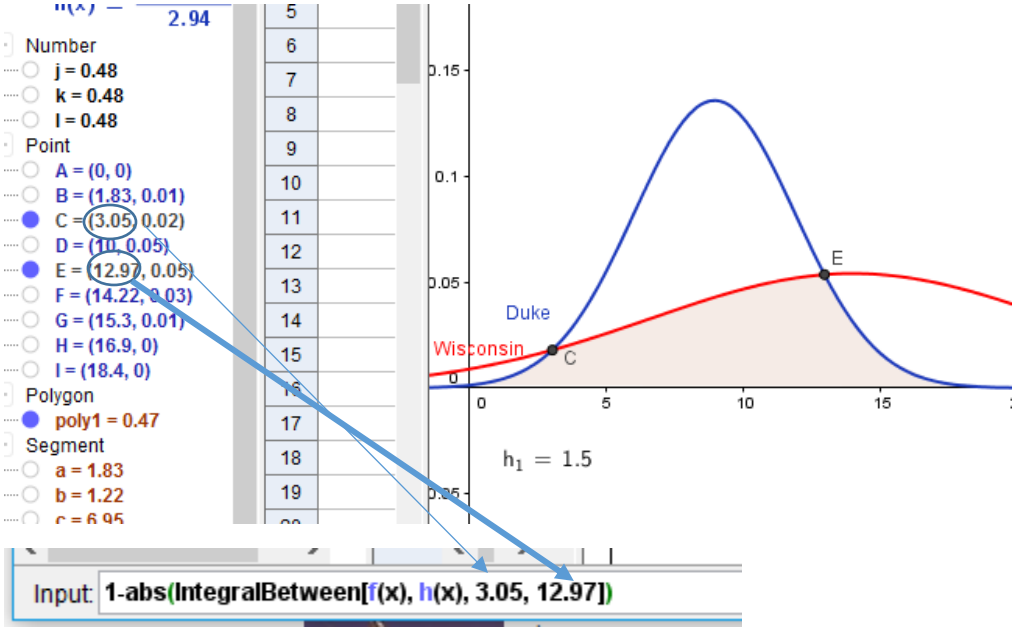
Appendix H

Select 2 points at the intersection of the 2 curves.

Select Intersect Tool



Use the x values as the Lower bound and Upper bound when computing the OVL:



Polygon area and Taking the Integral should match or be very close!

Example to try on your own: Select 3 cities to run a statistical analysis as we did above and fill in the chart below.

City	μ	σ	$P(X > 70)$	$P(50 < X < 70)$

The T test is one test used to help determine that the data could be from two different populations. Do the statistics from the different cities support that idea?

T-test Between Cities			
		Difference= P= t= SE= df=	Difference= P= t= SE= df=
	Difference= P= t= SE= df=		Difference= P= t= SE= df=

Draw the normal distribution curves for at least two selected cities. What is the probability the weather will be the same no matter which city you live in? (Find the Overlapping Coefficient, OVL)

Appendix I

Name _____

Does more education translate to higher salaries?

BUREAU OF LABOR STATISTICS

Earnings and unemployment rates by educational attainment, 2015

Education attained	Unemployment rate in 2015 (Percent)	Median weekly earnings in 2015
Doctoral degree	1.7	$1,623
Professional degree	1.5	1,730
Master's degree	2.4	1,341
Bachelor's degree	2.8	1,137
Associate's degree	3.8	798
Some college, no degree	5.0	738
High school diploma	5.4	678
Less than a high school diploma	8.0	493
All workers	4.3	860

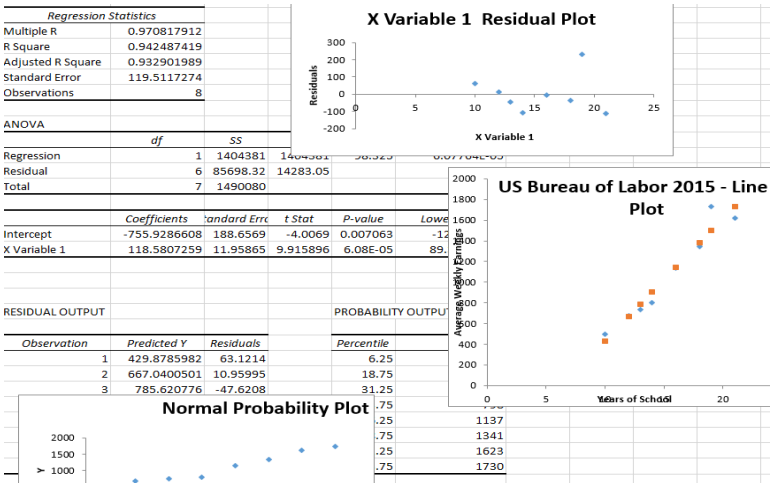
Note: Data are for persons age 25 and over. Earnings are for full-time wage and salary workers.
Source: Current Population Survey, U.S. Department of Labor, U.S. Bureau of Labor Statistics

Part I. Enter data on available technology and run a linear regression analysis. An Excel spreadsheet example is below. Technology steps are included on the last page.

Education Attainment	Number of Years in School*	Weekly Earnings
Less than high school	10	493
High School Diploma	12	678
Some college	13	738
Associate's Degree	14	798
Bachelor's Degree	16	1137
Master's Degree	18	1341
Professional Degree	19	1730
Doctoral	21	1623

***Please note that numerical values are assigned to the categorical data, Number of Years in School are assigned with Education Attainment.**

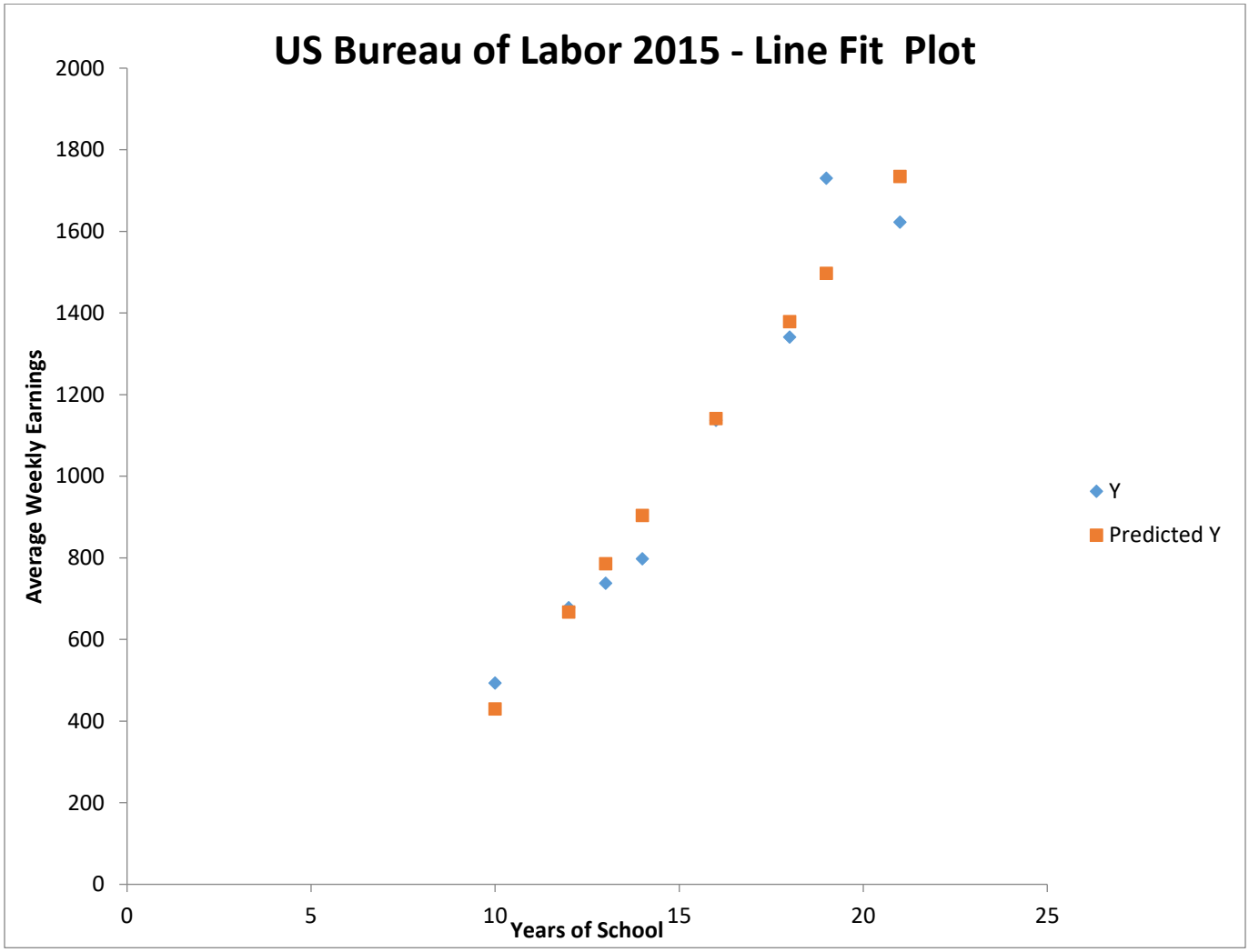
Data Analysis Results – What does this all mean?!



Part II: Correlation Coefficient, r.

Summary Output

Regression Statistics	
Multiple R	0.970817912
R Square	0.942487419
Adjusted R Square	0.932901989
Standard Error	119.5117274
Observations	8



Draw lines on the graph that represent the Correlation Coefficient, r.

What does r represent?

Appendix I

Write the formula for r.

Why is r squared and what is it called?

Given the r and $r - square$ value, what can be said about the data?

Give an example of an incorrect conclusion about the data.

Example: Use the following data to run a Regression Analysis each city. What are your observations regarding the data?

Month	Chicago Average High	Chicago Shootings 2016
January	32	300
February	36	190
March	46	310
April	59	305
May	70	405
June	81	425
July	84	425
August	82	500
September	75	350
October	63	410
November	48	395
December	36	80

Month	Boston Average High Temp (F)	Boston Shootings 2016
January	36	4
February	39	2
March	45	3
April	56	1
May	66	2
June	76	6
July	81	1
August	80	5
September	72	4
October	61	8
November	51	6
December	41	

Appendix J

Education Attainment	Number of Years in School	Weekly Earnings
Less than high school	10	493
High School Diploma	12	678
Some college	13	738
Associate's Degree	14	798
Bachelor's Degree	16	1137
Master's Degree	18	1341
Professional Degree	19	1730
Doctoral	21	1623

Education Attainment	Number of years in school	Weekly Earnings
Less than high school	10	493
High School Diploma	12	678
Some college	13	738
Associate's Degree	14	798
Bachelor's Degree	16	1137
Master's Degree	18	1341
Professional Degree	19	1730
Doctoral	21	1623

Appendix J

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.970817912
R Square	0.942487419
Adjusted R Square	0.932901989
Standard Error	119.5117274
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1404381.182	1404381.182	98.32499983	6.07764E-05
Residual	6	85698.3179	14283.05298		
Total	7	1490079.5			

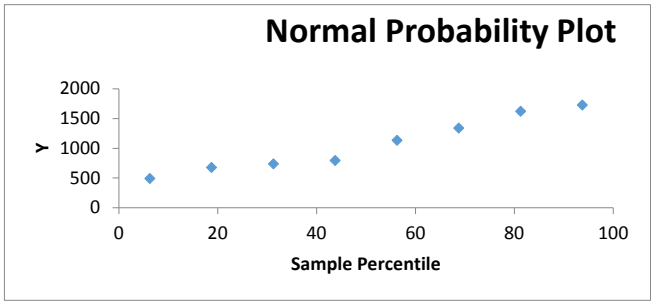
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Jpper 95.0%*
Intercept	-755.9286608	188.6569305	-4.006895792	0.007063304	-1217.55554	-294.3018	-1217.556	-294.3018
X Variable 1	118.5807259	11.95864923	9.91589632	6.07764E-05	89.31896538	147.8425	89.31897	147.8425

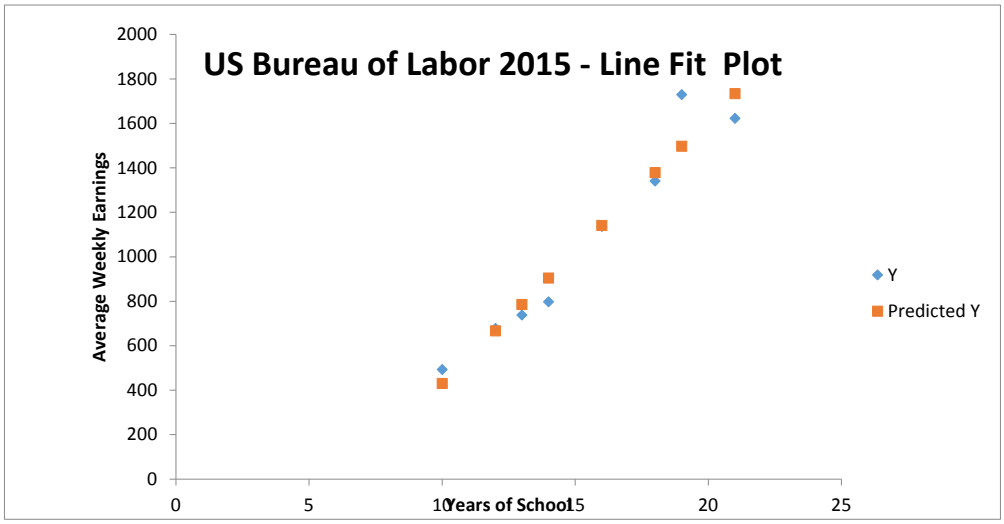
RESIDUAL OUTPUT

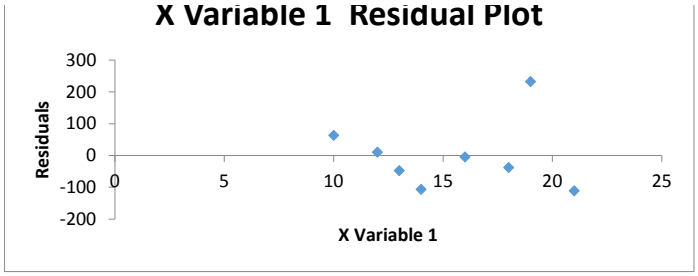
Observation	*Predicted Y*	*Residuals*
1	429.8785982	63.12140175
2	667.0400501	10.95994994
3	785.620776	-47.62077597
4	904.2015019	-106.2015019
5	1141.362954	-4.362953692
6	1378.524406	-37.52440551
7	1497.105131	232.8948686
8	1734.266583	-111.2665832

PROBABILITY OUTPUT

Percentile	*Y*
6.25	493
18.75	678
31.25	738
43.75	798
56.25	1137
68.75	1341
81.25	1623
93.75	1730






X Variable 1 Residual Plot
Residuals
300
200
100
0
-100
-200
0
5
10
15
20
25
X Variable 1