Danish Companies Dashboard: An Interactive, Geospatial Visualisation of Industries and Profit in Denmark

Tobias Hyrup, Pernille Matthews, David Nhan Thien Nguyen, Jakob Kusnick and Stefan Jänicke
Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark
E-mail: {tohyr17, pemat20, dangu17}@student.sdu.dk, {kusnick, stjaenicke}@imada.sdu.dk

***Abstract*—Profound knowledge of the business landscape is crucial for any company wanting to affect its position in the market. Whereas corresponding data is publicly available, visual interfaces that inform on the distribution of companies operating in different sectors are rare. To close the gap for the Danish market, we developed the Danish Company Dashboard (DCD), which uses the Danish Business Authority's database on company data to visually explore how the different companies, grouped by industries, are geographically scattered across Denmark on a regional and municipality plane. Moreover, the study and the accompanying visualisations provide insights into how the profit of each industry and company differs throughout the regions and municipalities, thereby supporting strategic decision making tasks of industry stakeholders.**

I. Introduction

With 5.8 million inhabitants per 2021 [23], Denmark is a rather small country, but it has proven a valuable global position with many major companies such as Maersk, Novo Nordisk, Ørsted, Vestas, Carlsberg, and many more. Not only does Denmark have highly succeeding large companies, but the country also has a thriving ecosystem for sole-man and entrepreneurial companies [2].

Denmark divides into five regions, each consisting of many different municipalities. Naturally, each region will attract specific industries due to its geographical characteristics.

Historically, there has been a negative discourse regarding the outskirts of Denmark. The expression "udkantsdanmark", translated to "rural Denmark", is a phenomenon referring to rural parts of Denmark dealing with high unemployment rates, poor education possibilities and more. The term gained popularity around 2010 [11], indicating that rurality continues to be a part of the Danish discourse. Rural areas and municipalities have attempted to make the discourse more positive [18]. Traditionally, the business environment in rural areas has primarily focused on agriculture, but the discourse change was meant to promote other industries in these areas. While the discourse proved largely unsuccessful, verified by the popularity of the term "rural Denmark", an investigation of the actual business environment in modern times can prove advantageous. Such an analysis can contribute to identifying opportunities and limitations and discerning the market in which each business resides and its surroundings.

LEVIA'22: Leipzig Symposium on Visualization in Applications 2022

In Denmark, all companies must register with the Danish Business Authorities. The registered companies are accessible using the digital portal called The Central Business Register, denoted as "Virk" [21]. Here it is possible to find data about companies such as CVR numbers, company addresses, industry type, financial reports, and much more. The CVR numbers given to each company are a unique identifier for companies in Denmark.

The need for accessing such company data can be many, including, but not limited to, business needs or personal interests. Despite this need, there is a noticeable lack of established and usable applications beyond paid services. Without such applications, investigating company data places a significant load on the users when processing the information computationally and cognitively. Subsequently, comparing companies across industries and geolocation becomes infeasible without extensive work.

We present the "Danish Company Dashboard" (DCD), which aims to bridge this gap between the available data and interpretability instead of relying on the current discourse-based assumptions. The DCD establishes a foundation for visual exploration by providing a geographical breakdown of the number and profit of companies in Denmark. It helps identify the geographical dominance across specific regions and municipalities while supplying additional information using supporting views. To realise this, we created cross-linked visualisations to form an interactive dashboard consisting of a map, scatterplot, and additional bar plots that supply stakeholders with detailed information on the Danish business landscape.

II. Related Work

Company data is accessible in Denmark using the Virk API. However, due to the resources required to use the API and then create visualisations and statistics upon that data, many companies who wish to know something specific about Danish companies look to other companies who sell this data in a ready-to-use format.

An example of such a company includes VAINU [20], a Danish company that provides an integration of their solution for companies that on a subscription basis can access information about any Danish company, supplied in a pre-processed

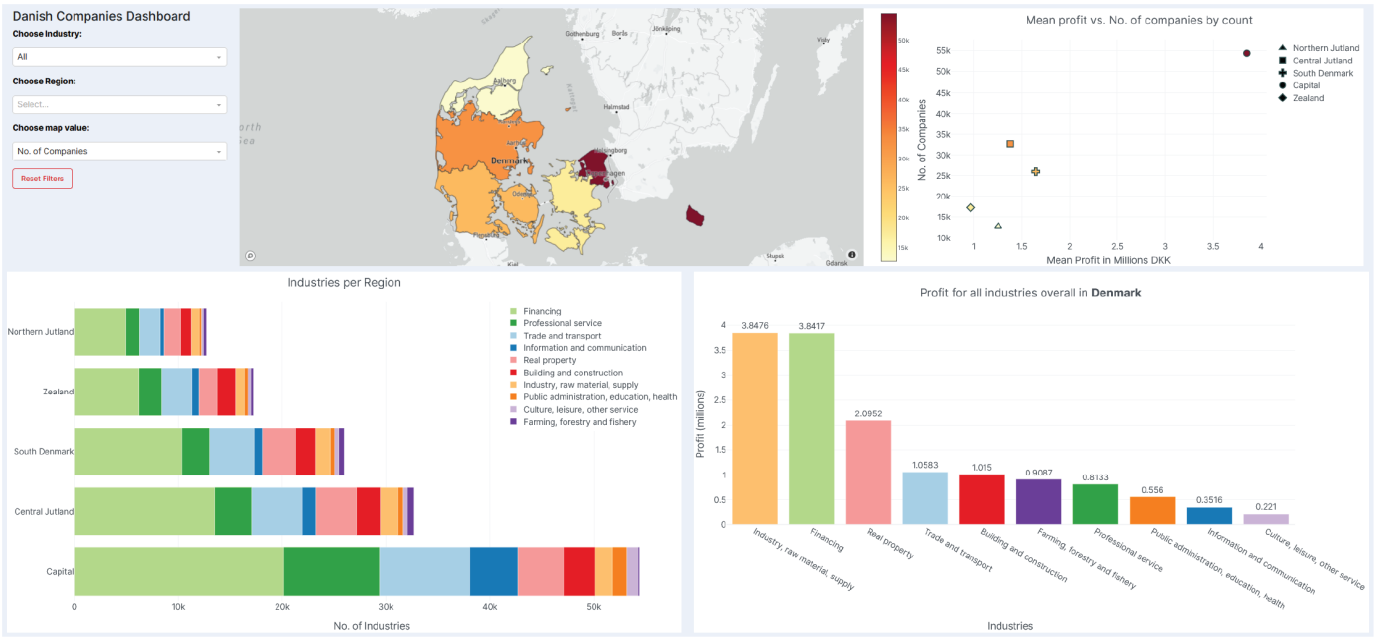

Fig. 1. The whole DCD in its initial state, where four juxtaposed views and the filtering system are visible.

and formatted manner. The result of companies similar to VAINU is that access to related work is restricted for this area as much of the work is behind closed doors.

Statistics Denmark is one of the few sources that have public work available regarding company data in Denmark. They collect, process, and publish statistics about the Danish society via their website [5]. Statistics Denmark has issued a statistics report utilising bar plots and tables to show the distribution of Danish companies across industries [4]. The DCD aims to extend aspects of this work and provide a different view on Danish companies by using juxtaposed views [12].

Only a few related works attend to visualising and exploring company-related data. Diamond and Mattia [6] attend to surveying the data visualisation tools commonly used in business-related contexts. Zheng [24] documents that such tools drive business performance as they yield ”effective strategic, tactical, and operational insights” for business intelligence. In a micro-study on four mining companies, MacDiarmid et al. [10] use charting tools to analyse the companies' performances. A similar approach to the DCD is given by Hausdorf et al. [8] who visualise the company register as a heat map to support exploring the local business community.

The DCD takes its main inspiration from the TreeeX system [9]. The work, in short, uses an interactive choropleth map as the primary view to analyse tree diversity and conservation status, both globally and nationally. Similar to the TreeeX system, the DCD focuses on supplying multifaceted analyses of the Danish company data using juxtaposed views, with the primary view being a thematic map, namely a choropleth map. While the DCD's only geographical focus is Denmark, it still has several levels as the data analysis is done both on a regional and a municipality plane.

III. Data Collection and Processing

The DCD can compare industries, geographical location, and profit by collecting data from the Danish Business Registry and scraping financial results. This data collection and processing procedure is an extensive and impractical procedure to perform by companies.

A. Data Collection

Determining which data is essential for the DCD is the initial step of the data collection. Virk's API enables searching on specific criteria, and for this study, the following applies:

- Companies that are active in 2020-2021.
- Companies with financial reports in XML format.

A crucial aspect of this study is that not all companies must submit a yearly financial report in Denmark [1]. The companies excluded from having to submit such a report are:

- One-person companies.
- Partnership companies where the responsible owner does not own a company with share capital.
- Companies that have gone bankrupt.
- Companies that are under reconstruction per bankruptcy law.

As this study requires a company to provide a financial report to gain insight into each company's profit, there is natural filtering of those companies not providing such a report.

B. Data Processing

In Denmark, each company must belong to at least one specific industry group. Typically, a company has one primary industry and many sub-industries attached to its identity. A total of 726 different industries exist in the Danish system, [3] but far from all are actively used. To promote statistical work and

simplicity to the number of industries in Denmark, Statistics Denmark encourages the use of ten primary industries [3], as seen below:

1) Farming, forestry, and fishery.
2) Industry, raw material, supply.
3) Building and construction.
4) Trade and transport.
5) Information and communication.
6) Financing.
7) Real Property.
8) Professional service.
9) Public administration, education, health.
10) Culture, leisure, other services.

These primary groups provide a generalised and simplified view of the industries in Denmark.

Certain complications occurred when accessing each company's XML file and scraping the profit. When extracting profit from each company's XML file, there was variation between currencies. Therefore, extracting each currency type was necessary to transform all non-Danish currencies into the Danish Krone. Out of 158,206 companies, the profit of 15,182 companies could not be scraped due to errors in access to the XML file, undetectable XML tag for profit and other unknown reasons. Excluding these companies from the final data set resulted in the total number of companies being 143,022.

IV. Visual Design

The Danish Companies Dashboard (DCD) consists of four interactive views. The upcoming subsections will explain each view and how they facilitate a comprehensive study of companies in Denmark. A key focus when constructing the DCD was to enable users of all backgrounds to use and interpret the dashboard. Therefore, emphasis is put on this focus when introducing each view. The creation of the dashboard uses the programming language Python with the open-source framework Dash [13] and interactive visualisation from the Plotly library [14]. The full DCD is seen in Figure 1.

A. Geospatial View of Companies

A geospatial view in current times has become a familiar visualisation due to the many worldwide COVID-19 visualisations [17], [22]. Danish news agencies have used geospatial maps of Denmark during most broadcasts with COVID-19. Thus, most Danes have been well-exposed to geospatial visualisations of Denmark [7], [19].

The primary view of the DCD is the choropleth due to its multi-functionality, depicted in Figure 2 through an example of the choropleth on two planes. The left view is the region-based representation of Denmark, which shows how the selected industry "Industry, raw materials, supply" is distributed by company count across Denmark per region. The right view is the municipality level; here, municipalities are shown. Providing a municipality level representation of Denmark enables more depth to the study. Accordingly, it is possible to pinpoint where the density of companies for all industries or within a particular industry is most considerable.

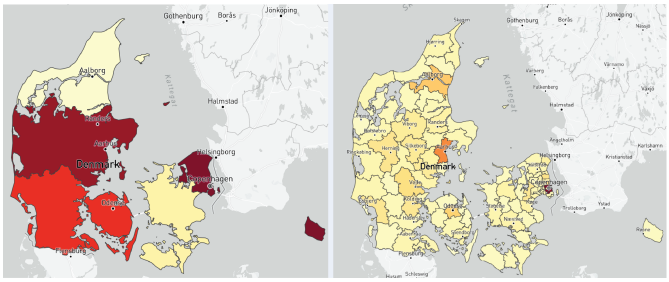

Fig. 2. The left view shows a choropleth of company count on a region plane within the industry: "Industry, raw materials, supply," while the right view shows it in the municipality plane.

B. Correlation Between Companies and Profit

Tightly connected to the choropleth view is the scatter plot view, seen in Figure 3. As a critical focus of this study is to investigate profit amongst companies and industries, the scatter plot provides valuable information on whether there is a correlation between the number of companies and mean profit for each region and municipality. Figure 3 shows two scatter plots, one for regions and the other for municipalities. By having mean profit and number of companies on the x and y axes, the correlation between the two can be investigated and compared for each industry across geographic positions. Regardless of whether the user understands statistical correlation, the scatter plot allows users to compare the values using the two axes. Moreover, each region is encoded as a separate symbol to allow comparison between regions.

C. Overview of Companies per Region

While the choropleth excels in showing a geospatial representation of companies, there is a lack of overview of the count of industries per region. There is much value behind knowing how many companies exist on a region or municipality plane. Correspondingly, gaining insight into each industry's proportion of the total number of companies in each region provides more in-depth knowledge.

Resultantly, a stacked bar plot, seen in Figure 1 bottom-left, is part of the DCD to convey this information. The view shows each region and the industry partitions that make up the national business environment.

An alternative to a stacked bar plot could be a radar chart [15, pp. 267-269] showing the industry distribution per region. However, such a view is unlikely to increase interpretability because of the many industries and the difficulty of comparing regions. Subsequently, a stacked bar plot provides the same information as a radar chart while simultaneously comparing all regions.

D. Industries per Municipality

The last view of the DCD is the horizontal bar plot, seen in Figure 4. The purpose of the horizontal bar plot is to provide information on the mean profit for each industry for the whole of Denmark, a particular region, or municipality. To

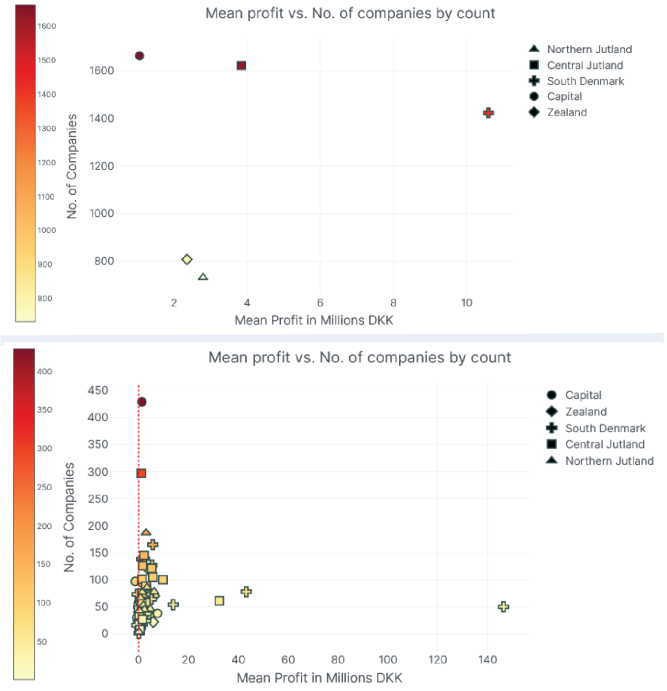

Fig. 3. Scatter plot view in two scenarios. The top view is the aggregated company count(x-axis) and company profit(y-axis) of "Industry, raw materials, supply" for all regions. The bottom view shows the same but on a municipality level for Denmark.

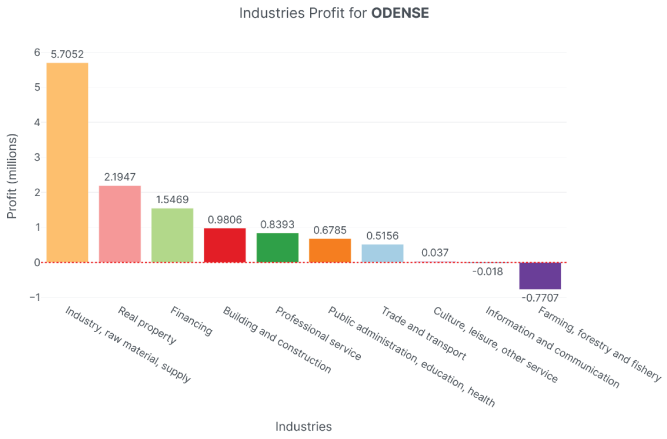

Fig. 4. Bar plot view showing the profit per industry for the whole of Denmark.

further enable distinguishment between industries, each bar is coloured uniquely. The bar plot is understood intuitively and proves valuable as it provides more specific insight into each municipality and its industries.

V. Interaction Design

The DCD is interactive and dynamically updates on particular actions to explore the provided data. This section will look into the intricacies of the DCD and the general workflow of using the dashboard. The aim of creating the DCD focuses on two primary use cases:

1) Identification of the most beneficial area to start a business within a particular industry.
2) Identification of what industry would be best to enter in a particular area.

A. Filters

The filtering system on the top left-hand side of the DCD, seen in Figure 5, enables the user to filter views according to specific inputs. The map, being the primary view, changes dynamically according to the filter inputs and the scatter plot is adapted correspondingly.

The three filters are as follows: 1) Choose all or a specific industry, 2) what regions should be selected, and 3) colour the map and scatter plot according to the number of companies or mean profit. An alternative method to selecting the filter values is by using the two bottom bar plots in Figure 1. By selecting an industry on the stacked bar plot, the filters change to match the given industry and region; the stacked bar plot does not change. Similarly, the regular bar plot filters by selecting one of the industry bars.

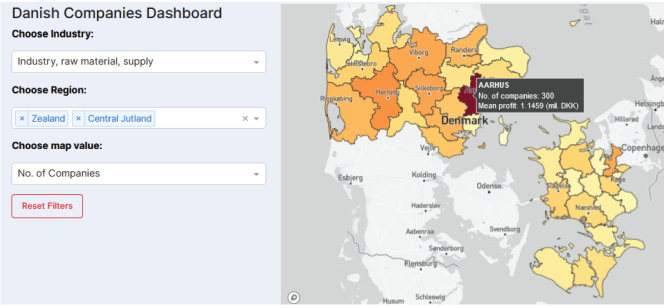

Fig. 5. Filters with two regions and the industry "Industry, raw materials, supply." applied and the resulting view of the map. Hovering shows additional information about a particular municipality.

B. Geospatial Interactions

The choropleth map's initial view shows the five Danish regions coloured by the number of companies, also seen in the complete DCD example in Figure 1. To get a more detailed view, the user can either 1) zoom in on the map, which will change the displayed areas from regions to municipalities, or 2) select a region by clicking on the map or selecting a region in the filter; selection via the filter allows the user to select multiple regions, see Figure 5. Additionally, the user can hover on any plot and get more information through to the textual tooltip.

In Figure 6, it is apparent that by using the scatter plot on the choropleth map's right-hand side, points are selectable using a selection tool. Once selected in the scatter plot view, the corresponding areas on the choropleth map are highlighted by lowering the opacity of all other areas. Subsequently, the user is provided with direct comparison of the choropleth map and

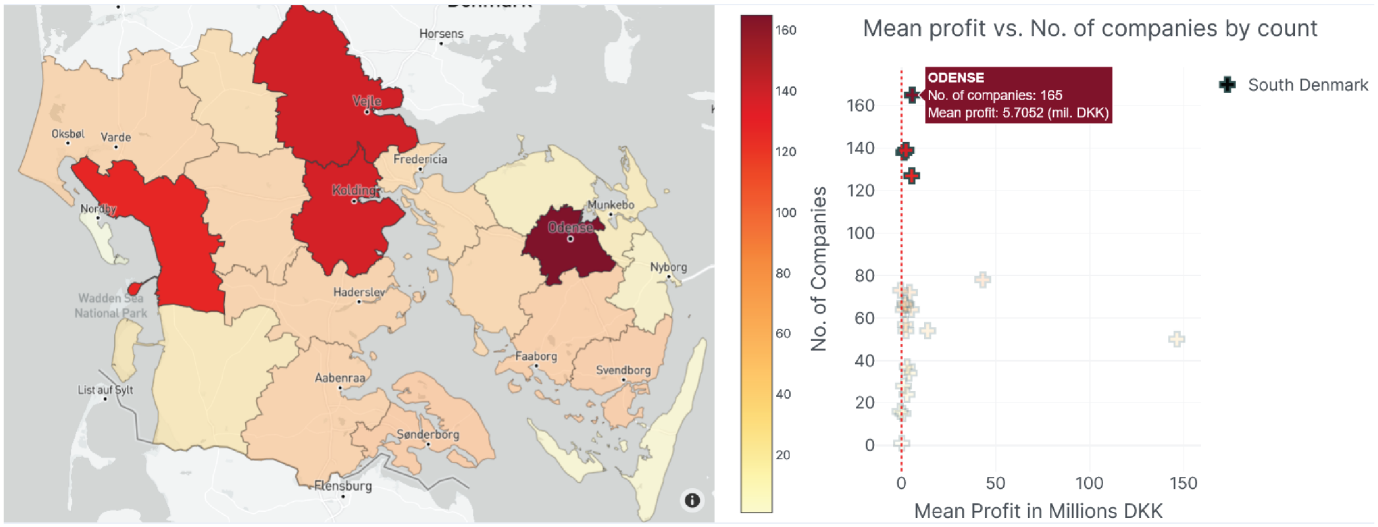

Fig. 6. Selecting points on the scatter plot highlights the corresponding areas on the map.

scatter plot, the highlighted areas, and the non-selected areas in one view.

C. Investigating Municipalities and Regions

Exploring a particular municipality or region can be done by selection on the map. As seen in Figure 4, the bar plot adapts to the selection by clicking on either a region or municipality. Similar dynamic behaviour is present when selecting a region in the stacked bar plot. This cross-filtering allows the user to use the choropleth map to find the desired area, investigate both company count and mean profits, and compare the profit within each industry for the given area.

D. Interactive Workflow

The interactive elements of the DCD allow the users easy access to relevant information given their use cases. The two primary use cases, as mentioned earlier, show how the visual design supports the workflow of real-life usage of the DCD given different contexts.

In **use case 1**, the user would filter for the given industry and zoom in on the map to reveal all municipalities. The most promising areas can be selected and highlighted on the choropleth map using the scatter plot. Then, the user can click on the individual municipalities and better understand how the industry compares to other industries in the area. Additionally, as seen below in Figure 7, the region that each municipality belongs to has a unique symbol that enables comparison on the scatter plot between regions on a municipality level. The information gained from this use case could form the foundation for finding a suitable area to start a business.

In **use case 2**, the user has a specific region or area in mind to start a company within. By clicking on either the given region on the map or specific industries and regions on the stacked bar plot, the choropleth map filters to reveal the given area exclusively. Additionally, the regular bar plot will adapt to show industry profit comparisons. It may also

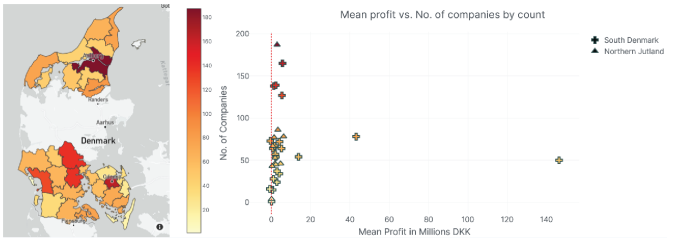

Fig. 7. Comparison of two regions on a municipality level for the industry "Industry, raw material, supply."

benefit the user to colour the map based on profit rather than count. Subsequently, the user can quickly get an overview of the competition in the area and make a decision based on that.

E. Implementation

The visualisations are made with the Python visualisation library Plotly, built on top of the JavaScript library D3.js and stack.gl [14]. Using Plotly, the graphs are rendered as JavaScript and can easily be made accessible online. The dashboard layout, interactivity, and cross-filtering are accomplished through the Python framework Dash created by Plotly [13].

VI. Discussion

The DCD presented in this paper provides an overview of companies in Denmark with a geospatial focus. The overview provides the user with an easy-to-understand look at the distribution of Danish companies and industries. This section discusses the rural Denmark phenomenon, current limitations, immediate user feedback, and future improvements for the DCD.

A. Comparison to Business Environment Assumptions

The discourse for rural areas has historically been centred around agriculture. Figure 8 shows a comparison of the industries "Farming, forestry and fishery" and 'Financing". It is

clear that the majority of the agriculture is centred around the western part of Denmark, "Jutland," in which many rural areas reside. Correspondingly, the 'Financing" industry is centred around the biggest cities in Denmark, namely Copenhagen, Aarhus, Aalborg and Odense. This differentiation between rural and urban areas follows the common discourse [11], [18] but does not show the whole picture. When investigating the individual municipalities, it is evident that the business composition in rural areas often favours industries such as "Industry, raw material, supply", "Trade and transport" and "Building and construction" in both mean profit and number of companies. The distinction can be made by looking at the bar plot presented in Figure 4 and through the choropleth map. Subsequently, the business environment does not follow the common discourse completely, which shows the benefits of the DCD.

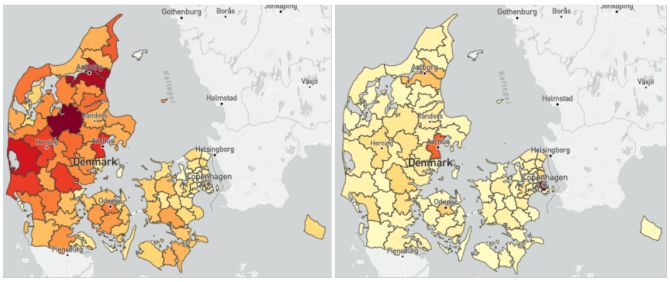

Fig. 8. Choropleth map on municipality level showing the number of companies in the industries "Farming, forestry and fishery" (left) and "Financing" (right).

B. Limitations

An apparent limitation of the DCD is that the overview is a static view of the current situation and, therefore, does not represent the temporal aspects presented by Statistics Denmark [4]. A temporal view provides information about how the industries are changing, providing valuable knowledge for companies. Temporal data would allow for additional plots such as a line plot over the historical data and the possibility of changing all the plots to represent each given year. Moreover, creating the possibility of a visualisation comparing year-to-year and, in turn, enable a month-to-month or seasonal comparison.

The geospatial view that the DCD provides, combined with scatter and bar plots supplements the graphs presented by Statistics Denmark. Furthermore, through cross-filtering and interactions, the possibilities in the DCD surpass the options provided by Statistics Denmark. Despite this, the DCD has significant limitations regarding the information one can extract from the dashboard. For example, it is evident that while the DCD presents an overview of Danish companies, it fails to provide details about specific companies and types of companies. This limitation affects the user outcome by creating several questions that the DCD cannot answer. Examples of such questions may include: "What companies dominate in this area?" or "How are the entry barriers for this location?".

C. Immediate User Feedback

Two small business owners, one with a web agency and another with a software development company, have provided immediate user feedback to the DCD, enabling an idea about real-world applications and possible improvements. The user feedback consisted of providing a short verbal description of the dashboard to provide context. After that, the user independently explored the dashboard using the Think Aloud Method [16, pp. 296-298]. The following were the results of each user.

Web-agency: The user quickly recognised possible applications of the DCD for their business to identify potential targets for doing business. Previously, they have had to manually sift through long excel sheets extracted from the Danish Business Registry. Immediately, they noticed that at least some of their objectives could be satisfied more easily by navigating the geospatial representations. A consideration made by the user was whether the ability to zoom in on individual zip codes might enable more profound insights. Also, the user noted that by identifying relevant areas, the user could perform additional research outside the dashboard.

Software Developer: The user's first consideration was how the DCD is usable in a real-world scenario. One use-case discovered was how it would be possible to utilise the overview and juxtaposed views to produce a marketing plan for their business clients. Additionally, the user thought the views would quickly identify interesting business areas and enable them to present that information to relevant stakeholders. The last consideration from the user is that the DCD allows identification of the most promising areas for advertising. In this association, the user suggested that information about demographics and employments rate for each region and municipality would be an insightful addition.

D. Future Work

The immediate user feedback provides a start for future work, and more user tests should follow moving forward to gain more significant insights into the development of the DCD. A greater variety of users should be selected to represent the user population better, and reoccurring results from such tests can be analysed, and the DCD adapted accordingly. The current test users are both from the industry "Information and Communication". Moving forward, future test users should be selected based on stratified sampling from all industries to ensure diversity and accurate representation of the user population.

New implementations should undergo an iterative process with the actual user to select appropriate modifications and implementations and ensure that real-world problems can be addressed by the DCD.

One possible improvement to the DCD is, as mentioned in the limitations, the addition of temporal data to represent changes in the business environment. Unfortunately, the API provided by The Central Business Register does not provide much historical data for financial reports. Consequently, the

supply of historical data from other sources is explicitly required.

The majority of the available financial data was successfully scraped, but some files were structured inconsistently. Future work could include further investigation into these anomalies and thereby include more data.

Lastly, updating the DCD automatically by creating a pipeline with new data would help keep the DCD relevant whenever used.

VII. Conclusion

The DCD is a visualisation tool used to investigate the count and profit of companies across Denmark on a region and municipality plane. DCD utilises a particular thematic map type, namely a choropleth. Upon filtering of the choropleth map, three supporting views change their appearance dynamically to supply the corresponding information of a user-provided filter. Though the DCD has limitations, it is relatively flexible with the current utilities that each user can use and supports multiple use cases.

By using geospatial data with juxtaposed views and interactivity, the DCD successfully extends the work by Statistics Denmark [5]. These extensions provide users with a more accessible comprehension of the structure of the Danish business environment. Furthermore, the immediate user feedback highlights the improvements the geospatial representations provide and places the DCD in direct real-world applications. Combined, the DCD bridges the gap between data and interpretability.

Additionally, the DCD creates an opportunity for much future work to extend the purpose and functionality of the current status. Much work is also available in the data collection and processing of the Virk API, enabling many exciting combinations of data to be processed and made available.

Lastly, the DCD exemplifies that assumptions based on social discourse may not produce an accurate representation of the Danish business environment. Subsequently, interpretable data representations are needed to supply a precise depiction usable for real-world applications. The use cases presented are examples of real-world applications that the DCD promotes.

Acknowledgements

The authors would like to thank Arvin Haghighi Aliari from Kolding Webbureau and Martin Wiboe from Substans Softwareudvikling ApS for their feedback on using the Danish Companies Dashboard.

Contributions

The paper is a result of a student project in which conceptualisation, data collection and processing, dashboard creation and the first paper draft were performed by Tobias Hyrup, Pernille Matthews, and David Nhan Thien Nguyen in equal parts. Jakob Kusnick and Stefan Jänicke supervised the project and revised the paper.

References

[1] Danish Business Authority. Hvilke virksomheder skal indsende årsrapporter - og hvem kan undtages? https://erhvervsstyrelsen.dk/hvilke-virksomheder-skal-indsende-aarsrapporter, Feb 2021.

[2] Danmark Statistik. De nye iværksætterselskaber vinder frem. https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=27743, 2016.

[3] Danmark Statistik. Statistisk Årbog. https://www.dst.dk/da/Statistik/nyheder-analyser-publ/Publikationer/VisPub?cid=22259, 2017.

[4] Danmark Statistik. Firmaer og koncerner. https://www.dst.dk/da/Statistik/emner/erhvervsliv/erhvervslivets-struktur/firmaer-og-koncerner, 2019.

[5] Danmark Statistik. Danmark statistik. https://www.dst.dk/da, 2021.

[6] M. Diamond and A. Mattia. Data visualization: An exploratory study into the software tools used by businesses. *Journal of Instructional Pedagogies*, 18, 2017.

[7] DR. Status på coronavirus lige nu. https://www.dr.dk/nyheder/indland/status-paa-coronavirus-lige-nu, 2022.

[8] A. Hausdorf, A. Niekler, and D. Wiegreffe. LocalCompanies: Visual Analytics of spatial aligned regional companies. In *LEVIA 2019: Leipzig Symposium on Visualization in Applications*, 2019.

[9] S. Jänicke, E. Beech, and M. Rivers. Exploring the diversity and conservation status of tree species with treeex. *Environmental Earth Sciences*, 78(21):1–10, 2019.

[10] J. MacDiarmid, T. Tholana, and C. Musingwini. Analysis of key value drivers for major mining companies for the period 2006–2015. *Resources Policy*, 09 2017.

[11] W. Malene Brandt and G. L. H. Svendsen. The rotten banana' fires back: The story of a danish discourse of inclusive rurality in the making. *Journal of Rural Studies*, 28:466–477, 2012.

[12] T. Munzner and E. Maguire. *Visualization Analysis & Design*. CRC Press, 2015.

[13] Plotly. Dash for python documentation. https://dash.plotly.com/introduction, 2022.

[14] Plotly. Plotly javascript graphing library plotly. https://plotly.com/javascript/, 2022.

[15] J. Schwabish. *Better data visualizations: A guide for scholars, researchers, and Wonks*. Columbia University Press, 2021.

[16] H. Sharp, J. Preece, and Y. Rogers. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 2019.

[17] Statens Serum Institut. Covid-19 - danmark (kommune) - opdateres hverdage kl 14. https://experience.arcgis.com/experience/aa41b29149f24e20a4007a0c4e13db1d, 2022.

[18] G. L. H. Svendsen. The right to development: construction of a non-agriculturalist discourse of rurality in denmark. *Journal of Rural Studies*, 20:79–94, 2004.

[19] TV2. Dagens coronatal: Stort overblik opdateret hver dag klokken 14. https://nyheder.tv2.dk/samfund/dagens-coronatal-overblik-over-smittede-indlagte-og-doede, 2022.

[20] VAINU. Cvr udtræk. https://www.vainu.com/da/cvr-udtraek-virksomhedsoplysninger/, 2022.

[21] Virk. Virk data. https://data.virk.dk/, 2021.

[22] World Health Organization. Who coronavirus (covid-19) dashboard. https://covid19.who.int/, 2022.

[23] Worldometer. Denmark population. https://www.worldometers.info/world-population/denmark-population/.

[24] J. G. Zheng. Data visualization in business intelligence. In *Global Business Intelligence*, pages 67–81. Routledge, 2017.