**DSC 205 – Spring 2024**

**Final Project**

**Guidelines and helpful hints**

* Final presentations will be held on 04.29 (in-person) and 05.01 (virtual). **Sign up for your presentation** [**here**](https://docs.google.com/document/d/13bEN_vP4lRR_kVCpwZtsCiK1vQNDtirfTn3sV4QA7KY/edit?usp=sharing) **no later than Friday 04.26.**
* Individual presentations are capped at 10 minutes, while team presentations are 15 minutes.
* For team presentation, all team members must participate.
* Create a PowerPoint / Google Slides presentation (4 slides total) that includes a title page (with team member names), a slide or two that describe the topic and the visualization goals you set out. It's important to provide your audience with a high-level view of your project before you demo your Streamlit app. After the demo, end your presentation with a slide that summarizes your main findings or observations.
* You're telling a story. Leave the audience with something to think about. The easiest way to lose your audience's interest is to read text from the slides or Streamlit app verbatim.
* Bring energy into your presentation. Your audience will not be interested in your project if you don't convey enthusiasm about your work.
* It's very easy to go over (or under) the allowed time. Rehearse your presentation and demo ahead of time. Move efficiently through your presentation, but don't rush.
* Presenters must have their cameras on for virtual presentations. If your computer doesn't have a webcam, sign up for a Monday slot.

**What to submit (by 11:59pm on 05.01 – hard deadline)**

* In Teams, submit the link to your GitHub repo (the same you used for building the Streamlit app in the cloud). The repo should contain:
  + Python code for your Streamlit app as well as any additional code you wrote for the project (pre-processing, etc.)
  + Copy or public links to datasets used in your Streamlit app.
  + PowerPoint / Google Slides presentation
  + One screenshot of your Streamlit app that includes your app's URL in Streamlit cloud.
* **A self-evaluation of your project**: Fill out the rubric below, compute your grade, and include comments that might help your instructor provide a more accurate evaluation of your work (e.g. particular challenges with the dataset that required far more time to pre-process the data than anticipated). Your evaluation will be taken into account when assigning final project grades.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **DSC 205 – Final project evaluation (Spring 2024). Final presentations Mon 04.29 and Wed 05.01 3:25-4:40pm.  10 minutes for individual presentations, 15 minutes for teams. Name:** | | | | | | |
|  | **Level of achievement** | | | |  | |
| **Criteria** | **Not evident (0)** | **Emerging (1)** | **Satisfactory (2)** | **Exemplary (3)** | **Score**  **0 to 3 (0.5-point increments allowed)** | **Weight** |
| **Data visualization goals and dataset selection** | Data visualizations questions vague and dataset not identified. | Data visualization questions **missing important details and focus** and/or dataset identified does not fully correspond to goals. | Data visualization goals are **mostly specified and focused** and dataset identified enables achievement of goals. | Data visualization goals **clear and compelling** and dataset identified enables achievement of goals. | **2.5** | **3** |
| **Data preparation** | No data preparation and cleaning performed | Data preparation and cleaning performed **minimal and/or not sufficient** to achieve project goals | Data preparation and cleaning performed **not trivial** and consistent with project goals | Data preparation and cleaning performed **requires significant effort** and is consistent with project goals | **3** | **3** |
| **Data visualizations** | Minimal or no data visualizations produced | Data visualizations produced **often** poorly labeled or **lack in variety and quantity** to achieve project goals | Data visualizations produced **generally** labeled and are **sufficient in variety and quantity** to achieve project goals | Data visualizations produced **consistently** labeled and are sufficient in variety and quantity to achieve project goals | **2.5** | **5** |
| **Streamlit application** | Streamlit application not developed | Streamlit application is not complete or does not include any interactive elements to enable the user to explore the dataset(s). **Application includes very limited descriptions of the visualizations and the project goals.** | Streamlit application is complete and includes some interactive elements to enable the user to explore the dataset(s).  **Application includes text explanations of the visualizations and the project goals.** | Streamlit application is complete and includes several interactive elements that engage the user with the exploration of the dataset(s). **Application includes text explanations of the visualizations and the project goals.** | **2.5** | **5** |
| **Streamlit cloud hosting** | Streamlit application not hosted on Streamlit cloud |  | Streamlit application hosted on Streamlit cloud |  | **3** | **2** |
| **Application of design principles (Gestalt principles, design guidelines, color maps)** | No application of visualization design guidelines and principles. | Minimal application of visualization design guidelines and principles. | Visualization design guidelines and principles used are adequate to communicate content clearly. | Application of visualization design guidelines and principles greatly enhance visual content. | **2** | **3** |
| **Visualization Walkthrough (4.22)** | Did not present visualizations. | Visualizations limited or under-developed. | Visualizations for some primary goals are developed. | Visualizations for all goals are developed, but missing some minor refinements. | **2** | **1** |
| **Final presentation:**  **quality of verbal and non-verbal communication** | Presentation not given | Speakers’ voices is consistently too weak or too strong  Speakers fail to use inflections to emphasize key points and create interest or speaker often uses inflections inappropriately Talking pace often too fast or too slow Transitions between speakers are somewhat choppy, confused | Speakers’ voices are generally steady, strong and clear  Speakers sometimes use inflections to emphasize key points and create interest Talking pace is generally appropriate Transitions between speakers are mostly smooth | Speakers’ voices are very confident, steady, strong, and clear  Speakers consistently uses inflections to emphasize key points or to create interest  Talking pace is consistently appropriate Transitions between speakers are seamless | 2 | **3** |
| **Final presentation (4.29 and 5.01)** | Presentation not given. | Final presentation provides **somewhat vague** description of project goals and how the produced visualizations and Streamlit application achieves the goals. Data story not compelling or not consistently supported by visualizations and presentation contents. | Final presentation provides **generally** **clear** description of project goals and how the produced visualizations and Streamlit application achieves the goals. Data story engaging and mostly supported by visualizations and presentation contents. | Final presentation provides **clear and compelling** description of project goals and how the produced visualizations and Streamlit application achieves the goals. Data story compelling and fully supported by visualizations and presentation contents. | **2.5** | **5** |
| **By 05.01:**  **Submit link to GitHub repo used for the Streamlit app in Teams.**  **(Code and datasets)** | No submission by due date.  **30-point deduction** |  |  |  |  |  |
| **Score (max = 88): \_74\_\_\_ Grade: \_\_A\_\_\_**  (Grading key: < 40: F, 40-49:D, 50-61: C, 62-73: B, 74-88: A)  Score is the sum of the products of individual scores with corresponding weights. Maximum score is the sum of the products of maximum scores with corresponding weights, excluding criteria marked as N/A.  **Comments:**  **The biggest challenge I ran into dealing with these data sets was the size of them. Coming into it I was optimistic about giving a world view, but quickly realized that was too much data for me to deal with, and it was almost impossible to display in an appealing way. I focused on 30 countries that I picked to try an encompass different parts of the world and big vs small in terms of land or population. While this was easier to work with, I do think it makes the maps look choppy and they aren’t as appealing to view as I would have hoped. I chose so many datasets to help me weave a story that had layers and transitions from one area of the topic to another, but it might have been better to stick with fewer and expand the amount of countries I used. Even only using 30 countries made some of the plots look crowded, so from a data stand point I don’t think I used a broad enough range but visually going at a global scale might have been too much. If I were to do it again I would probably focus on just the US, it would allow everything to look more uniform.**  **In the final update I had outlined 4 pages for my streamlit app, but once I saw everything together I could see that page 4 was much more dense than the others and that it would make more sense for page 2 to be dissolved and split between page 1 and former page 3. This kept the flow better of the stages of the story I was trying to create while also making each page more uniform. Also, in the final update I had written that I had two choropleth maps for deaths due to disasters and homelessness due to disasters and those worked well in colab but would not transfer over into streamlit so the visualizations had to be altered. I had originally planned to have two sets of columns, but the one on page 1 made the data seem too squished, so that was changed to provide each its own section.**  **I enjoyed using plotly to create animated charts, but I do think my some datasets might not have been ideal for them, the way I used them. The value ranges were sometimes so large that certain countries would overtake the others due to the axis scale having to be larger to accommodate one large value. From watching other presentations, especially Christian’s, I visually enjoyed all of his charts and could tell that he used plotly for all of them. If I were to redo the project, I think I would do the same thing to create more unity between all the data pieces. I wanted to try to incorporate plots of all different types, and I didn’t do much research into plotly besides the animated ones, so I was unaware of the static options that they had which I think would have provided me with a lot more options on how to display my data sets.**  **I always ran my code in real time on Colab first so it would be easier to troubleshoot errors step by step, and then I compiled them into the different pages for streamlit. All the plots ran perfectly individually in streamlit, but for some reason my plot of US CO2 emissions on page one and the plot of US precipitation anomalies on page two were affected by other aspects of my code. Before creating my cloud app I ran each page locally to see how it looked visually, and I realized that one page 1 before the user clicked the submit button every plot was shown correctly, but once the submit button was clicked it would alter the US CO2 plot even though that code was not nested within the limits of the button. I tried to troubleshoot the issue, but in the end ran out of time and therefore had to insert an image of what the chart should display as. I had a similar issue on page two with the precipitation plot and the death and homelessness column plots. Both were also being affected by submit buttons they were not nested in, and I attempted to split the page in half so those bottom plots would be on their own, only to realize that streamlit was displaying the precipitation anomaly plot three times, completely overwriting the death and homelessness plots. Originally there was a slider button above the two column plots to allow the user to select the year and it would display a seaborn bar plot with the countries on the y axis and the total death/homelessness numbers on the x axis, but I think since both of those were seaborn plots and the precipitation plot was also generated using seaborn that was causing the code to overlap. To fix the issue I also inserted an image of what the precipitation plot should look like and I got rid of the slider widget and replaced the two seaborn plots with plotly charts since I had not seen any overriding happen when there were multiple of that type on one page. My end result didn’t completely match what I had planned due to these unforeseen errors, and this situation is another reason why if I were to do it over I would probably create all of my visualizations with plotly- I believe it would have allowed me to keep more of my planned interactive elements and effectively run all of my code without having to insert images in their place.**  **I did not put much focus into color palette or gesault principles, even though I did try to make everything within the same color scheme. I feel I had so much going on already in terms of the amount of data, that I did not dive deeper into specific visualizations in the way that would have been necessary to display these principles.**  **Overall, I still love the concept that I created, but I ran into many problems along the way as well as had new ideas I wish I could have incorporated. I like to think I did the topic justice, but I know areas I would want to change and expand upon if I were to redo it.** | | | | | | |