Processing Book: Data Visualization Group A

1. Brainstorming and visual memos

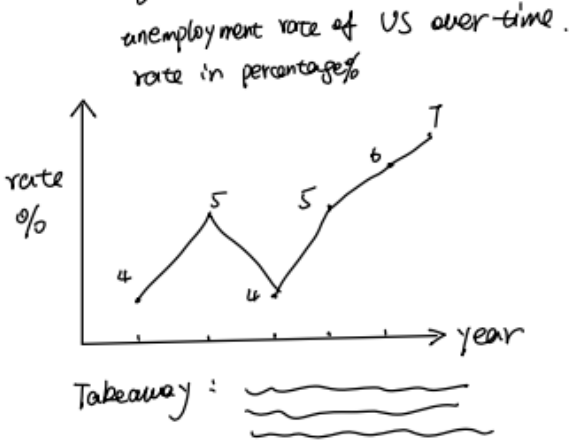
Unemployment rate

1. Dataset

① unemployment rate for the whole country.

semi-annual rate.

plot type: line chart.

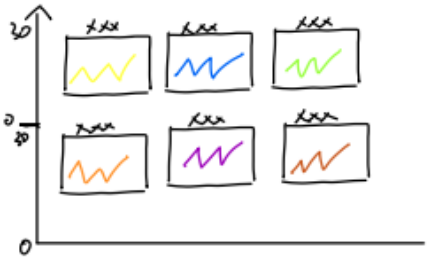


Takeaway:

② unemployment rate by industries

annual data

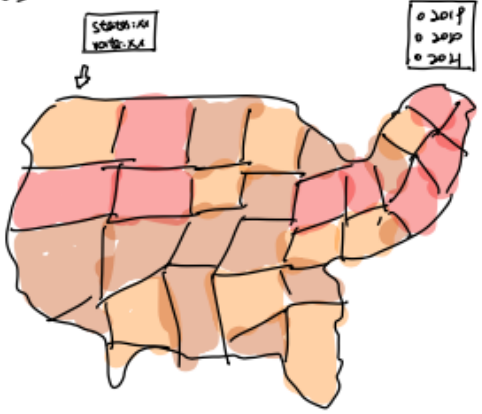
plot type: ~~facet~~ line chart

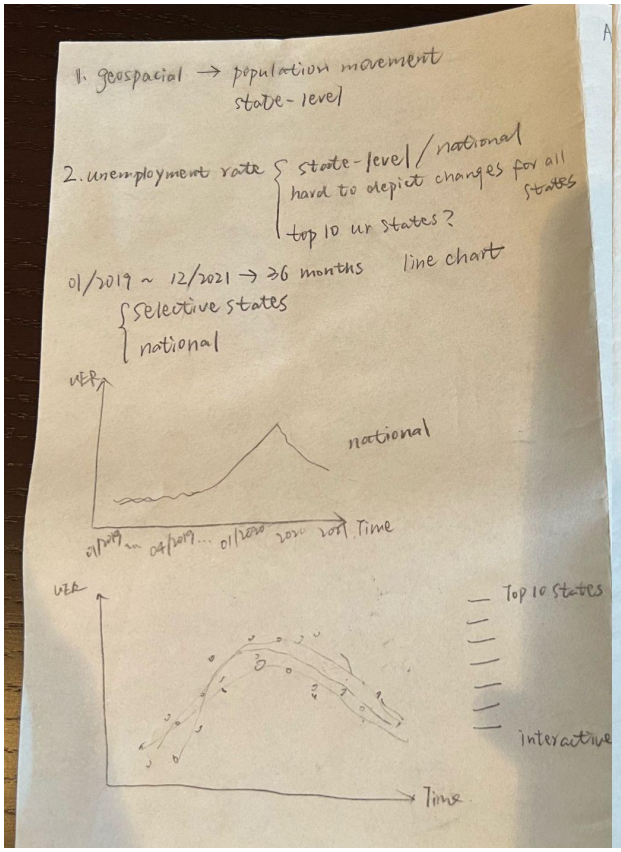


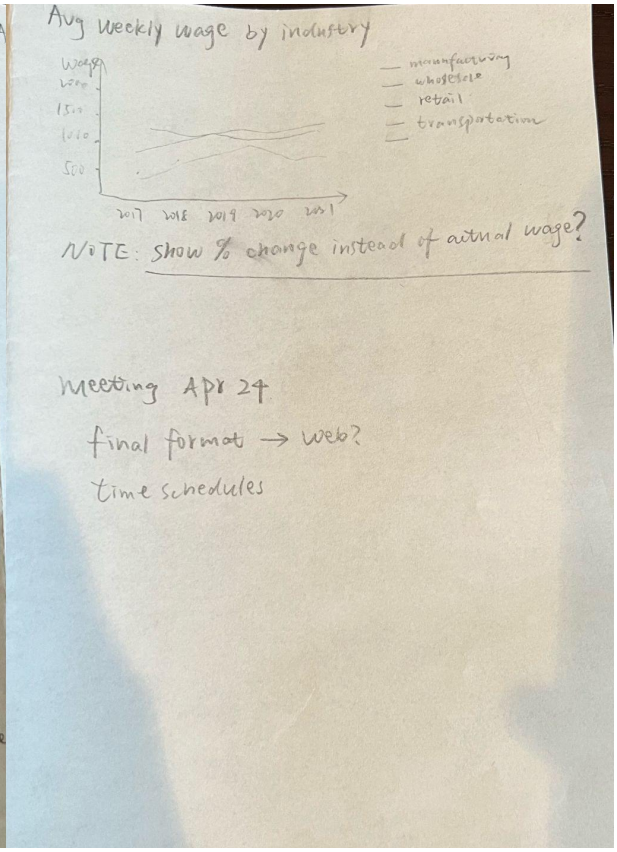
③ unemployment rate by US states

annual rate

plot type: map (interactive)







Dataset : US bureau of Labor statistics

timeframe

2005 ↔ interactive graph based on year → 2023

Map of US.

State | population group | employed | unemployed

↳ 16 ~ 24 ①
↳ 25 ~ 54 ②
↳ 55 + ③

(percent of population) (rate).

Making 2 interactive maps : employed vs unemployed.

population group ① → employed / unemployed
population group ② → " "
population group ③ → " "

1 map 2nd map.

Text Analysis

① learn to use API
② Visualizations → Word Cloud
→ positive / negative sentiment

2. **State trends in unemployment rate**
    a. Figure 1: A line chart to capture changes in average unemployment rates over time. We plan to first plot a national average, and also include separate line plots so as to contrast trends between different industries (static plot, with annotations indicating COVID-19 waves or lockdown periods)
    b. An interactive map showing all 50 states, with a timeline toggle, plus individual values for each state. Data ranging from Jan 2019 - Dec 2021, either monthly or quarterly values
    c. Data used: state level data

3. **Changes in wages over Different Industries**
    a. Weekly wage percent change over the year by industries (before and at the height of the COVID-19 pandemic)
    b. Datasets range from 2017 - 2021, with yearly records
    c. Line chart to show how wage and working hours have changed due to effects of the pandemic, and whether there are differences between industries.

4. **Differences in daily movement patterns of the population**
    a. Map-based visualization: since cities have gone into lockdown during the pandemic, people have switched to working from home and offices have closed down, their day-to-day movement ranges and patterns have drastically changed. We aim to visualize differences in people's movement patterns across time before and during the pandemic.

5. **Text Analysis**
    a. Step 1: we applied for the Twitter Development Platform account and obtained twitter API
        - Write a code to web scrape tweets for specific time intervals of use
        - Filter the location of tweets to the U.S. and only select tweets with the word "unemployment" in the text.
        - Extract 6 months worth of tweets for each analysis (2019, 2020, 2021), as 500 tweets are maximum we could download each time.
        - Time Frame: Download tweets for specific time intervals: Precovid (the first half of 2019), and the critical pandemic periods (2020 & 2021).
    b. Step 2: Data cleaning

- Preprocess the dataset by filtering out stop words, email addresses, and special characters.

c. Step 3: Visualizations:
   - Create word clouds for all three time frames to explore popular words in the tweets
   - Create a bar chart to compare the most frequent words in tweets of each year.
   - Perform sentiment analysis to understand people's opinions about the unemployment situation in the US in different time periods.
     - Lexicon to use: Bing (the Bing lexicon categorizes words in a binary method into positive and negative categories.)

6. Meeting Notes

- **First Meeting - Feb 28, 2022**

Initial Brainstorming Memo:

- How the COVID-19 pandemic has affected work - unemployment rate, work hours change, income change - trends across time
- Geospatial data for the US on state level https://gadm.org/download_country.html
- Map-based visualization: since people can work from home and offices have closed down, people's workplace can be flexible and some choose to move home or relocate to other places to work. We can visualize movement of people across time before and during the pandemic https://dataforgood.facebook.com/dfg/tools/movement-range-maps

—---

Datasets:

- COVID-19 Pandemic - Humanitarian Data Exchange https://data.humdata.org/event/covid-19
  - Data Explorer: https://data.humdata.org/visualization/covid19-humanitarian-operations/?layer=oxford_covid-19_government_stringency_index
  - WHO Global Covid-19 Cases Data
  - The New York Times Covid-19 Cases and Death in the United States (State Level Data)
  - Financial Times - Excess mortality during COVID-19 pandemic (https://data.humdata.org/dataset/financial-times-excess-mortality-during-covid-19-pandemic-data)
  - Government response stringency index:
    - Metrics: school closures; workplace closures; cancellation of public events; restrictions on public gatherings; closures of public transport; stay-at-home requirements; public information campaigns; restrictions on internal movements; and international travel controls.

        - A higher score indicates a stricter response (the strictness of government policies).
- WHO COVID-19 dashboard (https://covid19.who.int/)
    - Borrow visualization and interaction style
    - Any useful datasets available here? Can check
- Data For Good at Meta (https://dataforgood.facebook.com/dfg/covid-19)
    - **Population density maps**, movement range maps social connectedness (could be useful for network analysis)
    - Also has a great COVID-19 trends and impact survey dataset (for US and global levels)
    - Even has a tutorial on how to obtain, clean and visualize the dataset (https://www.youtube.com/watch?v=4Rfd7tGi9O0&list=PLITZ64izC0jyH1cDapSzDwJ1dx2ManGZW&index=7)
- US Bureau of statistics - Data on Labor

    https://covid-tracker.mckinsey.com/unemployment
    https://www.bls.gov/ces/tables/
    https://www.bls.gov/data/

- **Meeting Note from April 28, 2022**

Final format of deliverable: webpage

Project presentation websites due on May 4 at noon.
Student comments on other projects due on May 6 at noon.
Final student project, presentation video due on May 9.

Set up another meeting to display finalized visualizations, and unify design themes and plot styles. Then combine onto the same html page, with side panel to switch between pages.

Notes on individual parts:

Nicole:
- US national unemployment rate line graph: edit x-axis labels for year (Jan/June Year), and change the interactive label from “value” to “unemployment rate”.
- Unemployment rate by industry: edit facet size to display entire heading
- Map visualization of state unemployment rates for years 2019, 2020 and 2021: add color scale to create a choropleth map for state unemployment rates across the US, update the legend, and add percentage signs to the labels

Wenyi:
- Average weekly wage by industry: switch to percentage change, change title to “weekly wage (average quarterly) by industry, allow interactions to show multiple industries at one plot

Shauna:

- Remove jobs from list of tokens
- Download tweets from 2019, and not use 2022, let us know if you need help.
- Filter out tweets mentioning “trump” and do an individual sentiment analysis on those.

Toby:

- Show state and/or county level.
- Plot 1: movement range within state, US national map, average metric for all counties
- Plot 2: movement range across counties, allow selection to zoom/ focus on states, choropleth map
- Draggable axis toggle to allow visualization of changes across time.

- **Meeting Note from May 1st, 2022:**

Final format of deliverable: webpage

Project presentation websites due on May 4 at noon.
Student comments on other projects due on May 6 at noon.
Final student project, presentation video due on May 9.

- A general introduction and a conclusion
- Each one writes a one-paragraph introduction of the visualization
- Upload own process book onto drive
- https://mastering-shiny.org/action-graphics.html
- https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/
- Change all background color to white

Due Tuesday latest 2359:

- Final versions of .Rmd uploaded to google drive for compilation

Title: Impacts of the COVID-19 pandemic on the working population of the US
General introduction (wenyi):
Conclusion (everyone):

Individual introduction:
Nicole:
Wenyi:
Shauna:
Toby:

Next meeting: May 7, 4-6 pm

- Process book: draw and combine as pdf
- Presentation recording via zoom