

DSA 5200

Advanced Data Visualization

Final project

Final Project

- 2~4 students' teamwork, recommend to work in groups instead of individual
- CRN12841/12842: Presentation (10 points) + peer review report (15 points) + questions asking for the other teams' presentation (5 points) = 30 points
- CRN13473: Presentation (15 points, need to work on a more comprehensive topic) + peer review report for the in-class presentation record (15 points) = 30 points
- Students in CRN 13473 will record video for presentation
- Students in CRN 12481/12482 need to join in-person presentation

Presentation requirements CRN12841/12842

- Each team has 5 min to present
- All members in a team need to present within this 5 min
- Each team has 3 min to answer questions, limited to 5 questions to answer
- Don't start working on it now, I have more specific dashboard template requirements when we talk about the M15 Storytelling on 12/05/2025

Contents	credits
Background and introduction	1 points
Key questions to explore (3 types of questions)	1 points
Data cleaning, statistic and analysis	1 points
Dashboard to address the key questions	3 points (1 per key question)
Dashboard design	1 points
Summary and conclusion	1 points
Answer questions	2 points

Presentation requirements CRN13473

- 10-15 min to present, record a video and submit to BrightSpace
- All members in a team need to present within this 10-15 min
- Don't start working on it now, I have more specific dashboard template requirements when we talk about the M15 Storytelling on 12/05/2025

Contents	credits
Background and introduction	1 points
Key questions to explore (4 types of questions)	1 points
Data cleaning, statistic and analysis	3 points
Dashboard to address the key questions	8 points (2 per key question)
Dashboard design	1 points
Summary and conclusion	1 points

Peer review report requirements

- In-class presentation orders will be randomly generated at the beginning of the presentation date, therefore, peer review report order need to match with the presentation order. 1.5 point will lose for **each** of the mismatch order peer review comments.
- **Paper-based** peer review report for CRN12841/12842, submit one form per team. I will hand out this peer review report form at the beginning of the class and then collect them back at the end of the class.
- Peer review report for CRN13473 need to submit via BrightSpace.
- Peer review contents for each of the present team include 1) title and topic, 2) specific key questions asked by this present team, 3) data analysis methods used by this present team, 4) what type of plots this team used to address each specific questions, 5)* for CRN12841/12842, if you ask questions for this present team, what questions you ask, and what are their answers. 5)* for CRN 13473, select 5 present teams to comment what and where they can improve.

Questions asking for the other teams' presentation CRN12841/12842

- Each present team can only answer up to 5 questions, you can select any team raise hands that want to ask you questions.
- 1 question asking for the other present team earn 1 points, therefore if you want to have all the 5 points for questions asking, you must ask 5 questions. Only ask one question for one present team. You cannot ask more than one question for a specific present team.
- Tell your team number before you ask questions, so that I can have record for it to grade and check the contents in your peer review report.
- Questions can be, but are not limited to: data analysis methods, coding problems, dashboard design, visualizations for addressing the key questions point out to this topic, further directions, and so on.

Dataset selection

- I list the topics and dataset in the following slides
- Each dataset can be selected by two teams only (one team present on Tuesday 12/09/2025, and the other team present on Thursday 12/11/2025) — **first come, first served.**
- To select your dataset, please enter the names and school IDs of all team members in the following Excel sheet:
(https://docs.google.com/spreadsheets/d/1pPDZQ8L8ZeIq0xSlmDW1B9fJ4vMtWs5_/edit?usp=drive_link&ouid=107553936123392612833&rtpof=true&sd=true)

Topic for CRN 13473

❑ Recommendation dataset

- The following three datasets are collected from video-sharing mobile app. Based on the key questions you are interested; you can select one of them or subset of them. Follow the guide in the links for data download
- <https://kuairand.com/>
- <https://kuairec.com/>
- <https://kuisar.github.io/>

Following Topics are for CRN12841/12842

Topic 1: Health & Public Safety

❑ Dataset 1.1

- Provisional COVID-19 death counts and rates by month, jurisdiction of residence, and demographic characteristics
- https://data.cdc.gov/NCHS/Provisional-COVID-19-death-counts-and-rates-by-mon/yrur-wghw/about_data

Topic 1: Health & Public Safety

❑ Dataset 1.2

- Biennial Overview of Post-acute and Long-term Care in the United States: Data from the National Post-acute and Long-term Care Study
- https://data.cdc.gov/NCHS/Biennial-Overview-of-Post-acute-and-Long-term-Care/wibz-pb5q/about_data

Topic 1: Health & Public Safety

❑ Dataset 1.4

- Crime Data from 2020 to Present
- <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

Topic 1: Health & Public Safety

❑ Dataset 1.5

- Nutrition, Physical Activity, and Obesity - Behavioral Risk Factor Surveillance System
- <https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-behavioral-risk-factor-surveillance-system>

Topic 2: Finance & Market

❑ Dataset 2.2

- Small Area Health Insurance Estimates (SAHIE) using the American Community Survey (ACS)
- <https://www.census.gov/data/datasets/time-series/demo/sahie/estimates-ac.html>
- Only 2018-2022 data is enough for final project, no need for more years

Topic 2: Finance & Market

❑ Dataset 2.3

- Real Estate Sales 2001-2022 GL
- <https://catalog.data.gov/dataset/real-estate-sales-2001-2018>

Topic 3: Daily life

□ Dataset 3.1

- Crash Reporting - Drivers Data
- <https://catalog.data.gov/dataset/crash-reporting-drivers-data>

Topic 3: Daily life

❑ Dataset 3.2

- Consumer Complaint Database from CFPB
- <https://www.consumerfinance.gov/data-research/consumer-complaints/>
- The full database is huge, only 3 years data in MO is enough for the final project
- https://www.consumerfinance.gov/data-research/consumer-complaints/search/?dateRange=3y&date_received_max=2025-04-17&date_received_min=2022-04-17&page=1&searchField=all&size=25&sort=created_date_desc&state=MO&tab=List

Topic 3: Daily life

❑ Dataset 3.3

- Public School Characteristics 2022-23
- <https://catalog.data.gov/dataset/public-school-characteristics-2022-23-451db>
- See the following webpage to understand what each column value means
- <https://data-nces.opendata.arcgis.com/datasets/nces::public-school-characteristics-2022-23/explore?location=23.765896%2C-95.935812%2C3.48&showTable=true>

Topic 3: Daily life

❑ Dataset 3.4

- Rural and Small-Town America statistics
- <https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america>

Topic 4: Agriculture

❑ Dataset 4.1

- Crop Sequence and Nitrogen Fertilization Effects on Soil Properties in the Western Corn Belt
- <https://catalog.data.gov/dataset/data-from-crop-sequence-and-nitrogen-fertilization-effects-on-soil-properties-in-the-weste>

Topic 4: Agriculture

❑ Dataset 4.2

- Crop- and weather-dependent yield and wind erosion benefits from a conservation practices system
- <https://catalog.data.gov/dataset/data-from-crop-and-weather-dependent-yield-and-wind-erosion-benefits-from-a-conservation-p>

Dataset link

- https://drive.google.com/drive/folders/1mEdd4QdtHRZiPhFNLwjaL2K_bj8toQfR?usp=drive_link
- I downloaded the dataset from the web links I provided in each dataset slide. If you noticed something wrong, please let me know