

Project 2: Interactive Notebook + Infographic or Science Visualization

This is an overview of the project. See the links below for the deliverables.

This is a **group project** with several deliverables spread across the second half of the semester. In Project 1 you explored a dataset for yourself. In Project 2 you explain a dataset to someone else, in two forms:

1. An **interactive notebook** that guides a reader through your dataset and lets them explore your research questions, built with `ipywidgets` and `plotly`.
2. A **final piece** that distills your single most important finding into a standalone visual. Your group picks one track:
 - an **infographic** for a general audience, or
 - a **science visualization**: a publication-style figure for a specialist audience.

As well as a **final presentation** of your infographic/scivis.

The notebook is where the reader explores, and the final piece is for a reader who will never open your data; it has to make the point on its own.

Groups: 2–3 students. Choose your group in Canvas (under **People**) by **Wednesday, Oct. 21 at 10AM**; if you have no preference for group members, I will add you to a group by classtime.

Timeline

Deliverable	Due
Groups formed	Wed, Oct. 21
1. Project proposal	Sun, Oct. 25, 11:59pm
2. Narrative structure check-in	Sun, Nov. 15, 11:59pm
3. Final-piece sketches, in-class design critique	Fri, Nov. 20 (in class)
Critique reflection	Sun, Nov. 22, 11:59pm
Presentation slides	Tue, Dec. 1, 11:59pm
4. Final presentations	Dec. 2 – Dec. 4 (in class)
5. Final submission: interactive notebook + infographic / science visualization	Mon, Dec. 7, 11:59pm

Step 1: Data Selection and Proposal

Pick a topic and a dataset. You can use any data, as long as it isn't a trivial or toy dataset that has already been explored extensively (no built-in datasets like `penguins`, `mpg`, or `nycflights13`). If you aren't sure whether your dataset is appropriate, ask the instructor early.

Your **proposal** (no more than 1 page, one per group) should include:

- **Dataset:** where it comes from, who collected it and how, its size (rows × columns), and the key variables.
- **At least three research questions** you want to investigate.
- **Audience and track:** who you imagine reading your final piece, and whether you are leaning toward an infographic or a science visualization. You can switch tracks later, but tell the instructor if you do.
- **Planned interactions:** one or two ideas for what a reader might *do* in your notebook (filter by region, pick a year, compare two groups, ...) and why a static chart wouldn't be enough.
- **Team plan:** who is taking the lead on which part (data cleaning, notebook, final piece, slides). Everyone should contribute code and design work.

Step 2: Interactive Notebook (work on it throughout; final version due Mon, Dec. 7)

Start your notebook as soon as your proposal is in. Labs 7–9 (Plotly, ipywidgets, interactive trees) each give you a technique to try on your own data the same week. The notebook is where your group explores and **finds the main takeaway for your final piece**, and your results are what you'll use at the Nov. 15 check-in to decide what your final piece will be.

Your notebook should read like a short interactive article, not like a lab. Someone who hasn't taken this class should be able to follow it. It must contain at least two sections. Think of this as a newspaper article with interactivity.

About section

An introduction that catches the reader's attention and tells them what the project is about:

- What the dataset is, where it comes from, and any important limitations (missing values, sampling, what the data *can't* tell you).
- Your research questions and why a reader should care.
- A few key insights can be shown with **static visualizations** (matplotlib/seaborn is fine). Use the explanatory techniques from class: active titles, annotations, highlighting what matters, and direct labels.

Interactive section

At least three charts that meaningfully answer your research questions.

- **At least one `ipywidgets` control** (dropdown, slider, checkbox, toggle, ...) that changes what a chart shows.
- **At least one Plotly chart** with useful micro-interactions, such as custom hover text (formatted values and units, not raw column names), and zooming or legend toggling where it helps.
- **Every interactive chart** gets a 1–2 sentence note saying **what the interaction lets the reader do that a static chart wouldn't**. Interactivity that adds nothing (a dropdown with one sensible option, hover text that repeats the axis) should not be used. Think carefully about which charts actually benefit.
- Charts can stand alone, or widgets can control several charts at once (linked views). Linked views are not required but count toward creativity.
- Each chart is followed by **2–4 sentences of interpretation**: what pattern should the reader look for, and what does it mean?
- You may include static charts in this section as well; not every chart needs to be interactive. However, you should be more selective than in the first project: if you make a static chart and it doesn't add to your messaging, don't include it.

Takeaway section

End with a short summary of what you learned, and state the **single main takeaway** you've chosen for your final piece. This takeaway may be complex, but you should be able to write a summary sentence that will be used in the caption (science visualization) or title (infographic) of the main figure.

Technical requirements

`ipywidgets` need a running Python kernel, so **they will not work in an exported HTML file**. Because of that:

- Submit the `.ipynb`, your **data** (or a script that downloads it), and a `requirements.txt` (or a list of packages at the top of the notebook). If you don't use any packages beyond the ones we use in class, you don't need this file.
- Your notebook must run from top to bottom in a fresh kernel (**Kernel → Restart Kernel and Run All Cells**) with no errors and no manual steps.
- If your data is large (more than ~50 MB, e.g. Chicago Crimes, Yelp, Citi Bike), do your filtering and aggregation in a separate preparation notebook and load the smaller result in your main notebook. Submit both.
- Also submit an **HTML export** as a backup record. Plotly charts survive the export even though widgets won't. If your Plotly charts come out blank in the HTML, add `import plotly.io as pio; pio.renderers.default = "notebook"` at the top and re-export.

Step 3: Narrative Structure Check-in (Sun, Nov. 15)

By mid-November your notebook should have real results. Use them to decide what your final piece will be, and plan how it will tell its story before you start sketching. Submit a plan of **less than 1 page** (one per group):

- **What your notebook found:** two or three candidate findings from your notebook, each with the chart (screenshot or export) that supports it.
- **Your decision:** the **one main takeaway** you'll communicate (one sentence), your **track** (infographic or science visualization), and your **audience**. Briefly explain why you chose this finding over the others.
- **Your narrative structure:** which structure from class you'll use (for example, one of Segel & Heer's genres such as an annotated chart, partitioned poster, or magazine style; or a multi-panel $A \rightarrow B \rightarrow C$ figure for a science visualization), and why it fits your takeaway and audience.
- **A beat-by-beat outline:** what the reader sees first, second, and third. Cover the hook or context, the evidence (which charts, in what order), and the takeaway.
- **Chart plan:** which notebook charts you'll reuse, which you'll redesign for a static, explanatory version, and any new ones you need.

This plan is what you'll sketch from for the Nov. 20 critique.

Step 4: Design Sketches and Critique (Fri, Nov. 20; reflection due Sun, Nov. 22)

The critique is on your **final piece** (infographic or science visualization), not your notebook. Bring:

- **At least two different thumbnail layouts** for your final piece, with your one-sentence main takeaway written at the top. Build them from your narrative structure plan.
- **A rough draft of your strongest layout.** Paper, whiteboard photos, or quick slides are all fine. Charts can be placeholders or early versions from your notebook.
- **One sentence naming your audience**, so critics can judge whether the piece works for them.

Groups will rotate through critique stations giving and receiving feedback. Come with questions you want answered. By **Sun, Nov. 22**, submit a **critique reflection**: a photo or PDF of your sketches, the feedback you got, and what you plan to change in response.

Step 5: Final Presentation (slides due Tue, Dec. 1; presentations Dec. 2 & Dec. 4)

Each group has a **10-minute slot**. Present two things together:

1. **Your interactive notebook, live.** Don't scroll through every cell. Pick the interactions that show your most important insights and *use* them in front of the audience.
2. **A near-final draft of your final piece** (see Step 6). Feedback from this presentation should shape your revisions.

Every group member should speak. Give a brief overview of your project questions and data. Submit your slides by **Tue, Dec. 1** so all groups' slides are ready before presentations start. Feedback from the presentation can shape both your final notebook and your final piece.

Step 6: Final Submission (Mon, Dec. 7)

Your final submission is your revised **interactive notebook** (see Step 2) together with your **final piece**, described below.

Design a standalone visual that communicates your single main takeaway, combining text and charts. Choose a track:

Infographic (general audience)

- Built for a reader who has a minute or two and no background.
- A clear headline that states the takeaway, a visual hierarchy that tells the reader where to look first, and plain-language text.
- At least **two charts**, with consistent messaging.

Science visualization (specialist audience)

- A publication-style figure like you'd see in a journal: typically several panels labeled **A, B, C...**, with a **figure legend** (a title sentence followed by a description of each panel, sample sizes, and what any error bars mean).
- Follow the science-figure conventions from class: show the data rather than only summaries, use color carefully and accessibly, and keep fonts and alignment consistent across panels.
- At least **two panels** (A and B), with consistent messaging.

Requirements for both tracks

- **Final assembly can happen outside of code** (Figma, Canva, Illustrator, Inkscape, PowerPoint, ...). However, **every data-driven element must be generated from code** in your notebook: no hand-drawn, traced, or manually adjusted data marks.
- Text, layout, icons, and annotations can be added by hand.
- Cite your data source: on the graphic itself for an infographic, or in the figure legend for a science visualization.
- No deceptive encodings: check your piece against the ethics and deceptive-plots material from class.
- Submit as a **PNG or PDF** (vector PDF preferred for science figures).
- Include a **design rationale** (less than half a page): your audience, your takeaway, three key design decisions and why you made them, and **what you changed after the presentation feedback**.
- Include a short **contribution statement** listing what each group member did.

Grading

Project 2 is worth **20%** of your final grade. The project score is out of **250 points**, split across deliverables. The full rubric is at the end of this document.

Deliverable	Points
Proposal	25
Narrative structure check-in	25
Final-piece sketches + critique reflection	25
Final presentation	50
Final submission: interactive notebook (50) + infographic / science visualization (50) + design rationale (25)	125

Group members normally receive the same grade. If the contribution statement or instructor observation shows that work was very uneven, individual grades may be adjusted.

Recommended Data Sources

To get up and running quickly, consider one of these:

- [COVID-19 Worldwide Testing Data](#)
- [Citi Bike NYC trip data](#): large; pick a few months
- [Inside Airbnb](#): pick one or two cities
- [Top Spotify songs 2010–2019](#): ~600 Billboard top songs
- [Chicago Crimes, 2001–present](#): millions of rows; filter by year or type before loading
- [World Bank Indicators, 1960–2017](#): or browse the original data by indicator or country
- [Daily U.S. weather, 2017 \(NOAA GHCN\)](#): see the accompanying `weather.txt` for column descriptions
- [Opportunity Insights: social mobility in the U.S.](#): Raj Chetty's group; consider joining multiple files from the same paper
- [Yelp Open Dataset](#): large JSON; you don't need all of it (requires agreeing to Yelp's license)
- [Ookla Speedtest open data](#): mobile and fixed internet speeds

Science-visualization groups may want data from their own field, a lab they work in, or a public repository (e.g., NASA, NOAA, NCBI/GEO, the Allen Institute). Check with the instructor first.

Additional data sources

- [Google Dataset Search](#)
- [Kaggle Datasets](#)
- [Rdatasets](#)
- [UC Irvine ML Repository](#)
- [FiveThirtyEight datasets](#)
- [Awesome Public Datasets](#)

Submission Details

Submit everything on Canvas; each deliverable has its own submission page. One submission per group.

- **Proposal** (PDF): Sun, Oct. 25, 11:59pm
- **Narrative structure check-in** (PDF): Sun, Nov. 15, 11:59pm
- **Sketches** in class Fri, Nov. 20; **critique reflection** (PDF/photos): Sun, Nov. 22, 11:59pm
- **Presentation slides**: Tue, Dec. 1, 11:59pm
- **Final submission**: Mon, Dec. 7, 11:59pm
 - **Notebook**: `.ipynb` + data or download script + `requirements.txt` (if needed) + HTML export
 - **Final piece** (PNG/PDF) + **design rationale** + **contribution statement**

Generative AI use follows the course policy in the syllabus. Cite any AI help in your notebook and design rationale.

Acknowledgments

Components of this assignment were inspired by Adam Perer and Dominik Moritz's *Interactive Data Science* course at CMU.

Rubric

Interactive notebook + infographic or science visualization. 250 points total (20% of the course grade).
Rubric cells may not map exactly to specific point scores.

Deliverable	Points
1. Proposal	25
2. Narrative structure check-in	25
3. Final-piece sketches + critique reflection	25
4. Final presentation	50
5. Final submission: interactive notebook (50) + infographic / science visualization (50) + design rationale (25)	125

1. Proposal (25 pts)

Component	Excellent	Satisfactory	Poor
Dataset 7 pts	Source, collection method, size, and key variables are described; the dataset is substantial and not a toy dataset.	Dataset is appropriate, but the description is missing its source, size, or variables.	Dataset is trivial or built-in, or not described.
Research questions 8 pts	Three or more specific, answerable questions that go beyond single summary statistics.	Questions are reasonable but vague or shallow.	Fewer than three questions, or questions the data can't answer.
Audience, track & interactions 7 pts	Names a specific audience and track, and proposes interactions with a clear reason a static chart wouldn't be enough.	Audience or interactions named, but generic ("the public," "add a dropdown").	Missing audience, track, or interaction ideas.
Team plan 3 pts	Every member has a clear lead role.	Roles missing or unbalanced.	

2. Narrative structure check-in (25 pts)

Component	Excellent	Satisfactory	Poor
Evidence from the notebook 6 pts	Two or three candidate findings, each backed by a chart from the notebook.	Findings listed, but with weak or missing chart support.	No notebook results shown.

Component	Excellent	Satisfactory	Poor
Takeaway, track & audience 6 pts	A specific one-sentence takeaway, a chosen track, a named audience, and a clear reason this finding beat the others.	Decision made, but the takeaway is vague or the reasoning is missing.	No clear decision.
Narrative structure 6 pts	Names a structure from class and explains why it fits the takeaway and audience.	Structure named, but the fit isn't explained.	No structure identified.
Outline & chart plan 7 pts	Beat-by-beat outline (hook, evidence in order, takeaway) and a plan for which charts to reuse, redesign, or add.	Outline or chart plan present, but incomplete.	Missing.

3. Final-piece sketches + critique reflection (25 pts)

Component	Excellent	Satisfactory	Poor
Thumbnails & rough draft 10 pts	Two or more genuinely different layouts plus a rough draft of the strongest one, developed enough to critique.	Only one layout, layouts that differ only slightly, or no rough draft.	Missing or too rough to critique.
Takeaway & audience 7 pts	A specific one-sentence takeaway and a named audience, both reflected in the layouts.	Takeaway or audience stated, but vague or disconnected from the sketches.	Missing.
Critique participation & follow-up 8 pts	Gave specific, constructive feedback to other groups; submitted a critique reflection with a concrete plan for each piece of feedback.	Participated, but the feedback given or the follow-up plan was thin.	Absent, or no reflection submitted.

4. Final presentation (50 pts)

Component	Excellent	Satisfactory	Poor
Live demo 15 pts	Uses the interactions live to reveal the key insights; the audience sees why interactivity matters.	Shows the notebook, but scrolls through cells rather than using the interactions purposefully.	No demo, or the notebook breaks during the presentation.
Final piece draft 13 pts	A near-final draft whose takeaway and design choices are explained.	An early draft, or shown without explanation.	Missing.

Component	Excellent	Satisfactory	Poor
Structure & time 11 pts	Clear context → evidence → takeaway; finishes within 10 minutes.	Generally organized, but ran over or rushed.	Disorganized, or significantly over time.
Delivery & team 11 pts	Every member speaks; confident and audience-facing.	Uneven speaking roles, or delivery read from the screen.	One person presents everything, or hard to follow.

5. Final submission (125 pts)

5a. Interactive notebook (50 pts)

Component	Excellent	Satisfactory	Poor
About section 8 pts	Engaging introduction to the data, its source and limitations, and the research questions; static charts use explanatory techniques (active titles, annotation, highlighting).	Data and questions introduced, but limitations are missing or the static charts are exploratory-style with no emphasis.	Missing or reads like lab notes.
Answers to research questions 10 pts	At least three charts meaningfully answer the research questions, and each is followed by an interpretation of what to notice and what it means.	Charts address the questions, but interpretations just describe the chart rather than explain it.	Fewer than three charts, or charts unrelated to the questions.
Interaction design 12 pts	Uses ipywidgets controls and Plotly micro-interactions, each with a clear purpose, stated in a note on what it lets the reader do that a static chart couldn't. Linked views or other thoughtful touches earn full marks.	Requirements met, but some interactions add little (redundant hover text, a dropdown with no meaningful choices).	Missing ipywidgets or Plotly, or interactions that get in the way of reading the chart.
Visual design 8 pts	Effective chart types and encodings; clear titles, labels, and units; formatted hover text; accessible color.	Mostly effective, with a few encoding or labeling errors.	Poorly chosen chart types or encodings; unlabeled axes; raw column names.
Narrative & takeaway 6 pts	Reads like a short article a newcomer could follow; ends with a clear main takeaway.	Mostly followable, but the transitions or the takeaway are weak.	No narrative thread; no takeaway.
Reproducibility 6 pts	Runs top-to-bottom in a fresh kernel with no errors; data (or a download script), requirements, and HTML backup included.	Runs after minor fixes (a path, a missing package).	Does not run, or data is missing.

5b. Infographic / science visualization + rationale (75 pts)

Component	Excellent	Satisfactory	Poor
Main takeaway 10 pts	A single, clear takeaway a reader gets within seconds; stands on its own without the notebook.	Takeaway present, but competing with secondary points or requiring effort to find.	No clear takeaway; reads like a collage of charts.
Audience fit & genre 10 pts	Infographic: plain language and a headline a general reader understands. Science vis: labeled panels and a complete figure legend (sample sizes, error bars defined) at journal standard.	Mostly fits the audience, but uses jargon (infographic) or has an incomplete legend (science vis).	Wrong level for the audience, or ignores the conventions of the genre.
Composition & visual design 25 pts	Clear visual hierarchy and reading order; consistent alignment, type, and color; effective encodings; accessible color.	Generally well designed, with some clutter, misalignment, or competing emphasis.	Cluttered or confusing layout; ineffective encodings.
Data integrity 5 pts	Data marks generated from code; source cited (on the graphic for an infographic, in the figure legend for a science visualization); no deceptive encodings; uncertainty shown where relevant.	Minor issues (source hard to find, a questionable axis choice).	Misleading encodings, hand-drawn data marks, or no source.
Design rationale & revision 25 pts	Rationale explains audience, takeaway, and three design decisions; the revision clearly responds to presentation feedback; contribution statement included.	Rationale present but thin, or the revisions are minor.	No rationale, or no evidence of revision.