

Data Viz Hackathon

Jeremy Mikecz

Goal

Students will be divided into teams. Each team will be assigned one of the following datasets and topics:

Group	Dataset	Topic
	538 Bechdel dataset	gender representation in movies
	RCPA3 States dataset	crime
	RCPA3 nes (National Economic Survey) dataset	explaining Americans' happiness with life
	RCPA3 nes (National Economic Survey) dataset	News consumption
	538 college_all_ages dataset	employment and earnings by major
	Titanic dataset	Who survived the Titanic?

With the time allowed, each team is to create a series of visualizations with their scaffolding (title, axes labels, legends, labels, annotations, captions, etc.). Some recommended steps:

1. identify the data variables you will use
2. map the data variables to visual variables and choose plot types to represent your data
3. add in highlights, callouts, scaffolding, text, etc. to help highlight patterns and tell stories with your data
- 4.

After completion of the visualizations, all students will vote for the winning team (but can't vote for their own team) in the following categories:

1. most truthful / functional
2. most insightful / enlightening
3. most beautiful / eye-catching

Rules

1. you may reuse code you find online (i.e. from R Graph Gallery) to help you build the components of your visualization (but must design and arrange the components yourself)
2. you may apply code generated by a Large Language Model to build the components of your visualization - but all ideas and the organizations of these components should be your own.
3. all students and the T.A. may vote (but cannot vote for their own team)

Setup

```
library(tidyverse)

Warning: package 'tidyverse' was built under R version 4.4.3
Warning: package 'tibble' was built under R version 4.4.3
Warning: package 'tidyr' was built under R version 4.4.3
Warning: package 'readr' was built under R version 4.4.3
Warning: package 'dplyr' was built under R version 4.4.3
Warning: package 'stringr' was built under R version 4.4.3
Warning: package 'forcats' was built under R version 4.4.3

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
vforcats    1.0.0     v stringr   1.5.1
v ggplot2   4.0.0     v tibble    3.2.1
v lubridate 1.9.3     v tidyr    1.3.1
v purrr    1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

1. Import Datasets

1.1 Import RCPA3 NES & States datasets

```
# install.packages("RCPA3")
library(RCPA3)

Warning: package 'RCPA3' was built under R version 4.4.3
data(package = "RCPA3")

# head(nes)

# head(states)
```

1.2 Import 538 datasets

```
# install.packages("fivethirtyeight")
library(fivethirtyeight)

Warning: package 'fivethirtyeight' was built under R version 4.4.3
Some larger datasets need to be installed separately, like senators and
house_district_forecast. To install these, we recommend you install the
fivethirtyeightdata package by running:
install.packages('fivethirtyeightdata', repos =
  'https://fivethirtyeightdata.github.io/drat/', type = 'source')
data(package = "fivethirtyeight")

# head(bechdel)

# head(college_all_ages)
```

1.3 Titanic dataset

```
titanic <- as_tibble(Titanic)
```