

The in-class midterm has two components. The first section (20 minutes) consists of short answer questions focusing on concepts and basic vocabulary. You are not allowed to use your notes, computer, books, the internet, or any other source of assistance. The second section (40 minutes) is a short programming activity similar to the exercises we have been doing in class and for assignments. You are welcome to use a computer, your notes, the course handouts, completed activities, the internet, the book, or any other sources, but are not allowed to work with or communicate with anyone else.

Two main topics

- Data visualization
- Data wrangling

Important skills and knowledge

- Understand the layered grammar of graphics; answer the following questions about a graphic:
 - What variables constitute the frame?
 - What glyphs are used?
 - What are the aesthetics for those glyphs?
 - Which variable is mapped to each aesthetic?
 - Which variable, if any, is used for faceting?
 - Which scales are displayed with a guide?
 - What raw data would be required for this graphic, and what form should it be in?
- Understand properties of effective visualizations
 - Storytelling/context: What story does the graphic tell? Does it reveal information that is latent in the data?
 - Ethics: Does the graphic mislead the audience? Is the data source clear?
 - Design: Visual perceptions of aesthetics such as position, length, angle, direction, shape, area, volume, shade, color
- Identify quantitative vs. categorical variables
- Identify the unit of observation (case) of a data table and which combination of variables uniquely define each case
- Understand the difference between three broad families of functions in data wrangling:
 - Reduction functions: take a variable as input and return a scalar (e.g., mean, sum, n, min, max, median, n_distinct)
 - Transformation functions: take one or more existing variables as input and return a new variable (e.g., age < 21, age >= 65, age == 39, age %in% c(18,19,20), log10(population), weight/height)
 - Data verbs: take an existing table as input and return a new table (e.g., group_by, summarise, mutate, select, filter, arrange, pivot_wider, pivot_longer, left_join, inner_join, semi_join, anti_join)
- Determine which data wrangling steps are required to rearrange a data table in a way that is conducive to making visualizations (i.e., glyph-ready format), generating new tables of a specific form, or finding specific summary statistics (e.g., the mean value of a variable for a particular group of observations)
- Find code online and adapt it to your given task

Programming skills

- Generate and knit an R Markdown file to ensure reproducible work
- Load data into R with a function such as `read_csv`
- Basic R skills (e.g., assign a variable, find the dimensions of a table, access a column of a table, find the names of variables)
- Use the `ggplot` package to create univariate, bivariate, and multivariate visualizations to compare different variables (both categorical and quantitative)
- Visualize spatial data as scatterplots, contour maps, or choropleths
- Change the variables (columns) of a data table with `select` or `mutate`
- Change the rows of a data table with `filter` or `arrange`
- Perform grouped summaries with `group_by` and `summarise`
- Change the case/format of a data table with `pivot_wider` and `pivot_longer`
- Join two tables of data together with either a mutating join (`left_join`, `inner_join`, `full_join`) or filtering join (`semi_join`, `anti_join`)
- Use the pipe notation to execute multiple wrangling operations at once
- Manipulate dates with the `lubridate` package