CEM1002

Neil Montgomery

2015-09-21

A little bit about R

- Assignment: <- (RStudio shortcut Alt + -)
  - = works but should be reserved for setting parameters in function calls
- Most basic object type: vector or elements
- Numeric, character, date, boolean
- Possible to have matrices, but not R's strengh
- Other object types: lists and data frames
- Accessing list and matrix elements may be useful for us when writing a report that requires extracting a particular result from some statistical analysis.
- **Data frames** are the most important for us, by far. All our datasets will be data.frames.

Special kind of vector: `factor`

- What `R` uses to store a categorical variable.
- But, can lead to challenges with data import and export.

Useful R packages (examples to follow)

- `rio` data import/export made (often easier)
- `lubridate` for dates and times when present in data
- `dplyr` for nice data manipulation
  - `https://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html`
  - `https://rpubs.com/justmarkham/dplyr-tutorial` (with link to related video)
- `ggplot2` for nice graphics
  - `http://ggplot2.org/`
  - 'ggplot2' book available on SpringerLink
  - `http://www.cookbook-r.com/Graphs/`

"...and then..."

- `dplyr` and `ggplot2` let you manipulate and plot data in a way which, at first, will seem bizarre, but in my opinion follow the way people actually think about each step.
- "Take the data, and then focus on specific rows, and then..., and then..."
- "Decide on an x and a y variable, and then make a scatterplot..."

Univariate summaries (last week)

- Categorical variable: *table*
- Numerical variable
  - Location: mean, median, percentile (quantile)
  - Spread: range, variance, standard deviation
- "Observed sample . . . "

Graphical

- Categorical: barplot (`R note: "factor"`)
- Numerical: histogram, density plot, **boxplot**