Introduction to 

Sociological Research Methods

Jason Thomas, IPR Research Scientist
thomas.3912@osu.edu
https://www.asc.ohio-state.edu/thomas.3912/

Oct. 19 & 21

Agenda

- Tuesday (any questions)
  - basic setup & syntax
  - finding help
  - data types & structures
  - working with data
- **Thursday**
  - statistical analyses & exporting results
  - noteworthy R packages for plotting & data carpentry
  - dynamic documents
- GitHub: https://github.com/buckipr/r_intro_srm

Statistical Analyses

Work with a script for a while (getting a little into programming)

Noteworth Packages

Welcome to the `tidyverse`

- `dplyr package` – "grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges"
- `ggplot2 package` – "ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details."

DPLYR

- **Logic:** "By constraining your options, it helps you think about your data manipulation challenges."
  - R **for Data Science**: https://r4ds.had.co.nz/
- Grammar for **Rows** – selecting and organizing observations/cases
- Grammar for **Columns** – cleaning and creating new variables
- **Merging Data**
- Additional features of the language:
  - tibble data structure (similar to data frames, but crankier)
  - "pipe" %>% for stringing multiple commands together
  - useful for keeping number of objects to a minimum and for plotting (e.g., adding separate symbols for subgroups)

Grammar for Rows

- `filter()`
- `slice()`
- `arrange()`
- *Groups of rows*: `summarize()`
  - useful companion: group_by()
  - collapse across rows with `summarize()`

Grammar for Columns

- `select()`
- `rename()`
- `mutate()`

GGPLOT2

- *Logic:* Create a graphic through a sequence of **Layers**
  - Data
  - Mapping - how variable are related to the plot
  - Statistical Transformation - add statistical summary to plot
  - Geometric Object - type of plot (scatter, time series, box-plot)
  - Positional Adjustment - e.g., adjust the placement bars on a plot

GGPLOT2: Resources

- `ggplot2` is a member of tidyverse: https://ggplot2.tidyverse.org/
  - cheatsheet
- Hadley Wickham's book: ggplot2: Elegant Graphics for Data Analysis
  - introduction to grammar of graphics: Section 1.2
- Useful chaptes from R for Data Science
  - Chapter 3: Data Visualization
  - Chapter 7: Exploratory Data Analysis
- Addition references for data visualization
  - Kieran Healy's Data Visualization: A Practical Introdcution
- Galleries
  - r-graph-gallery
  - extensions ggplot2

Dynamic Documents

- Use a *single* document to weave empirical results and substantive text
  - reduce the number of steps to create the final document and efficiently reproduce those steps
- There are several tools for creating dynamic documents that are available in different stats packages
- `R Markdown` is a popular R package (and rightfully so)
  - output options include: pdf (paper and slides); MS Office (Word & Powerpoint)

Markdown

Syntax for adding styling to your documents

- ****this is bold**** and ****this is italicized****
- lists with asterisks, dashes, plus sign (nested lists with indentation)
- section headers (add more # for subsections)
  - `# Section 1`
  - `## Section 2`
- Web links: `[link name](link url)`:
  - e.g. `[Rmarkdown](https://rmarkdown.rstudio.com/)`
  - Rmarkdown

Installing R Markdown (& dependencies)

Install the following programs (all of the default options are fine)

- (good idea to go with RStudio for this)
- Mac users will install **X Quartz** program from https://www.xquartz.org/, which is used by Mac to show plots
- LaTeX is needed to create PDFs (can do this through R packages as we will now see)

R Packages for R Markdown

- Open R Studio, then copy and paste the following command into the **Console** pane.
- *Note*: this assumes you do not have LaTeX installed on your computer (which is needed to create PDF files with R Markdown).

```
install.packages(c("rmarkdown", "tinytex", "rprojroot", "Rcpp"),
                 repos = "http://cran.r-project.org")
```

R Packages (cont.)

- Install a small version of LaTeX (using the tinytex package). This will take a few minutes...
- Windows users may see a warning about needing to install the Rtools program, and 2 warnings complaining about missing files associated with `luatex.dll`
  - but these can all be ignored.

```
tinytex::install_tinytex()
```

R Packages (cont.)

- If you would like to test your `tinytex` installation, close R Studio, then re-open it and run the following command

```
tinytex:::is_tinytex()  ## should return TRUE
```

- Note the 3 colons after `tinytex`. If the above command returns a value of `TRUE`, then you are all set.

R Packages (cont.) Warning Messages

- Here are 3 screenshots of the informational and (2) warning messages that Windows users can safely ignore.

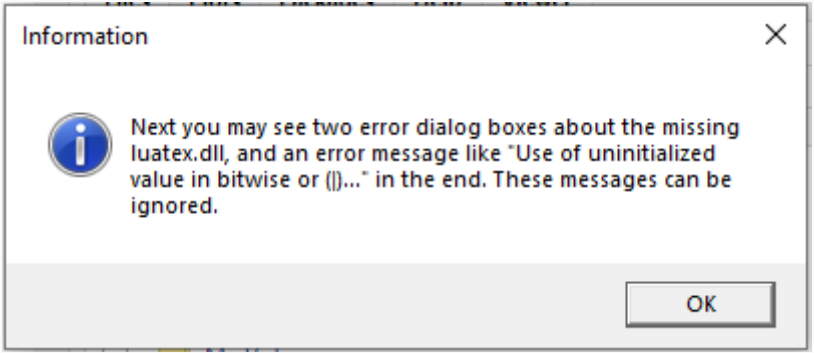

Figure 1: informational message

R Packages (cont.) Warning Messages 1

(can safely ignore)

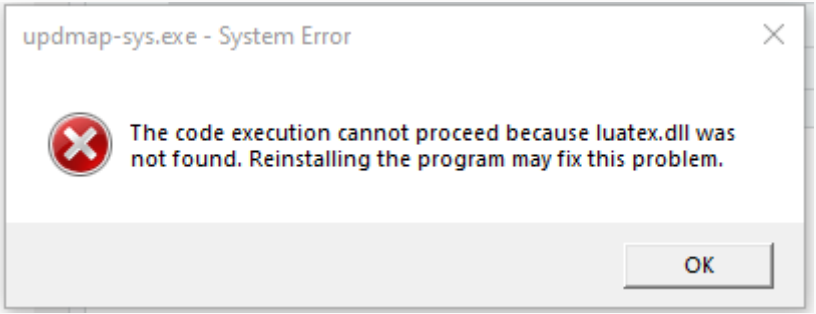

Figure 2: 1st warning message

R Packages (cont.) Warning Messages 2

(can safely ignore)

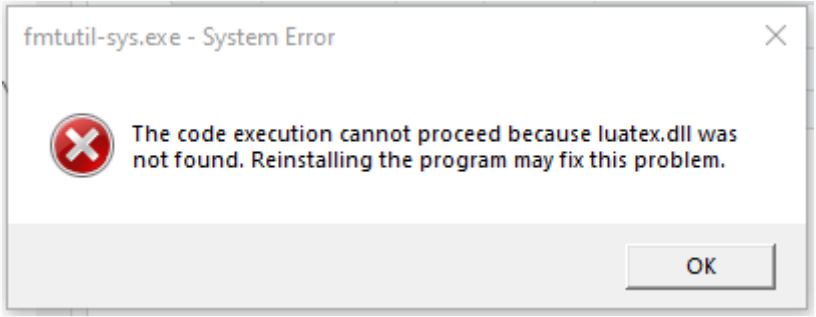

Figure 3: 2nd warning message