

Intro to R

Elizabeth Schoen Simmons

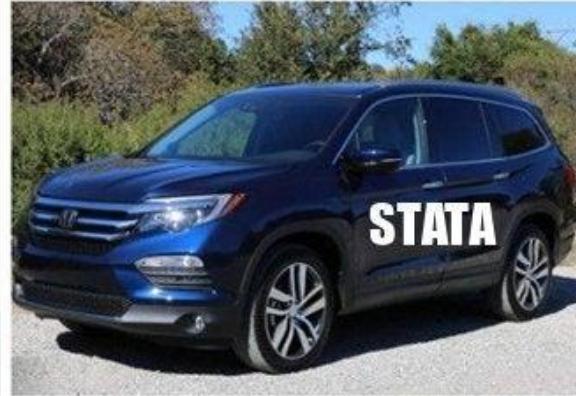
Sacred Heart University

February 2023

R u ready?

- PollEv.com/elizabethsimmons903

If statistics programs/languages were cars...



Why R?

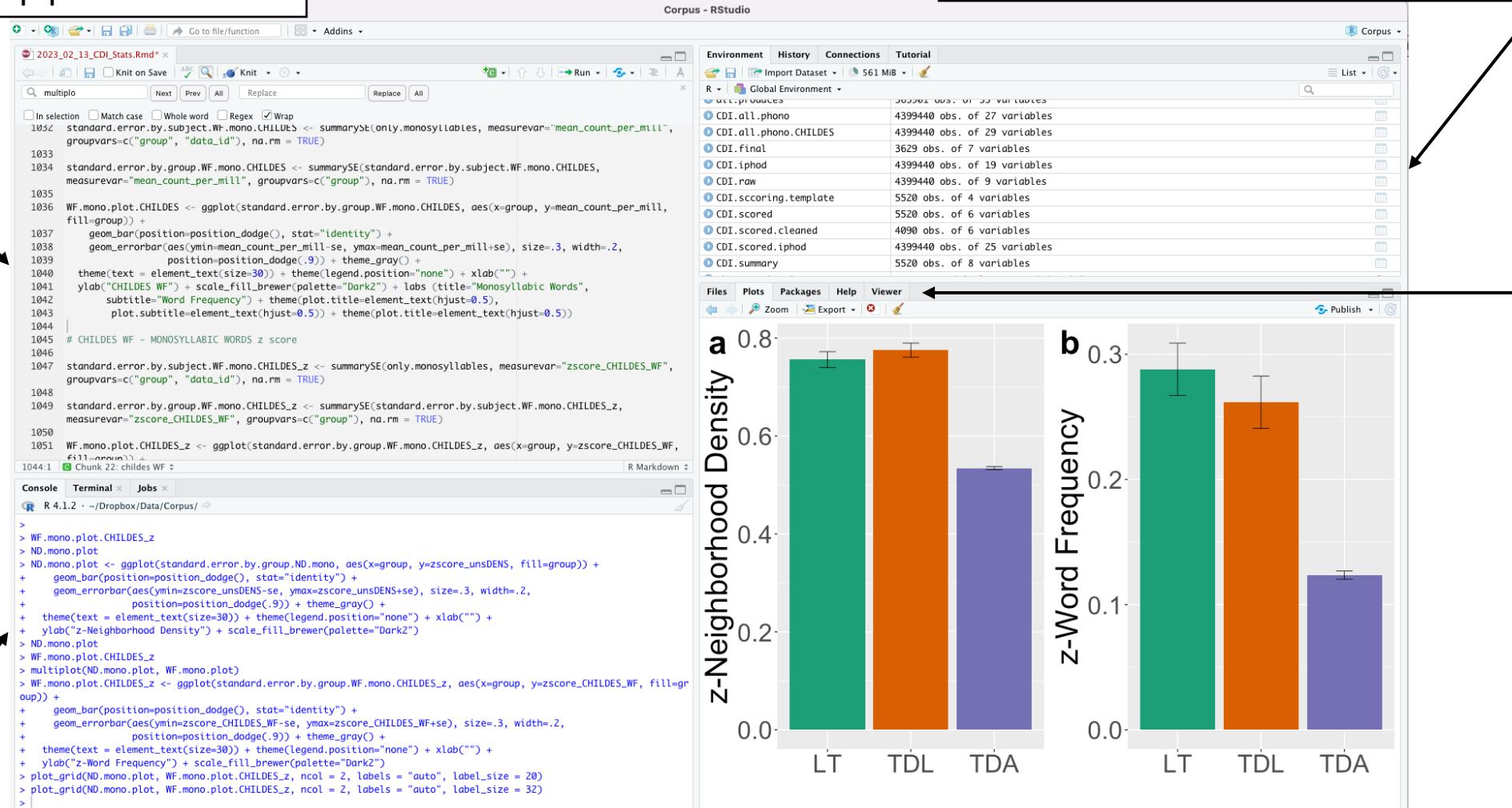
- Reproducibility!
- Transparency!
- Efficiency!



The basics

- **R** is a coding language
- You can use **RStudio**, a coding environment, to run R code
 - Write and execute code without using the R terminal
 - Helps with transparency
 - Easy to run long scripts and short chunks of code
- **R packages** are free, mini-programs that allow you to complete a range of functions

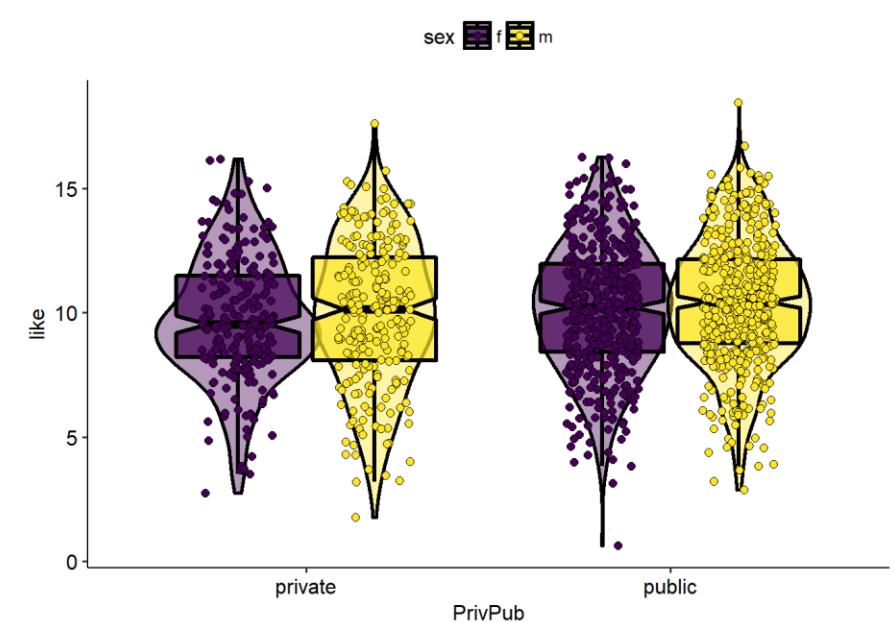
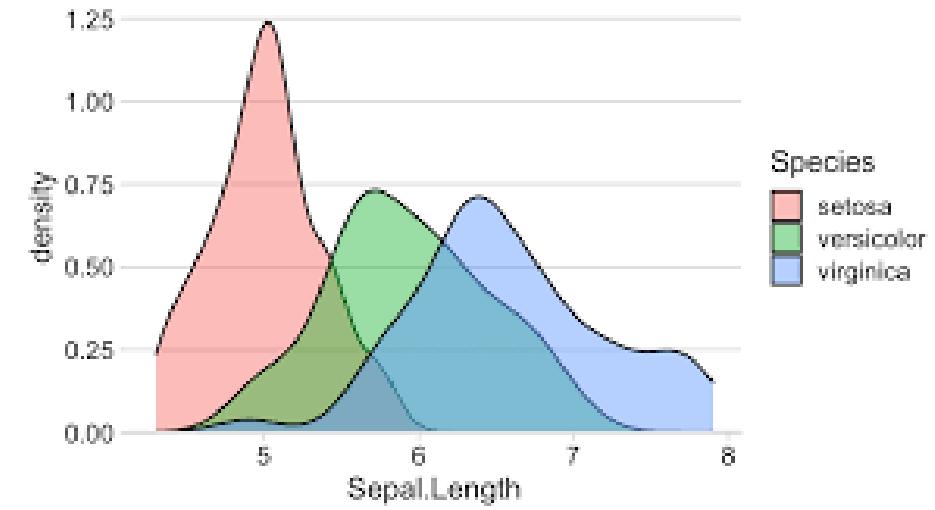
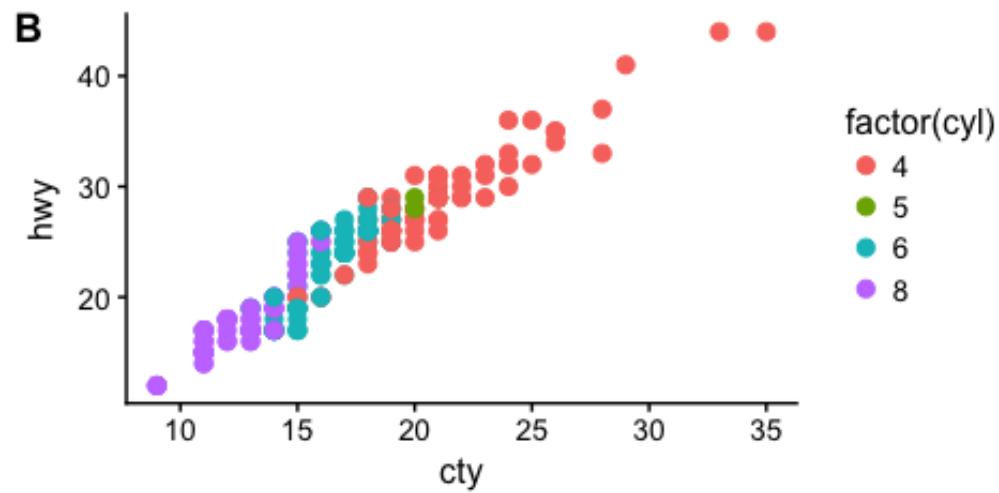
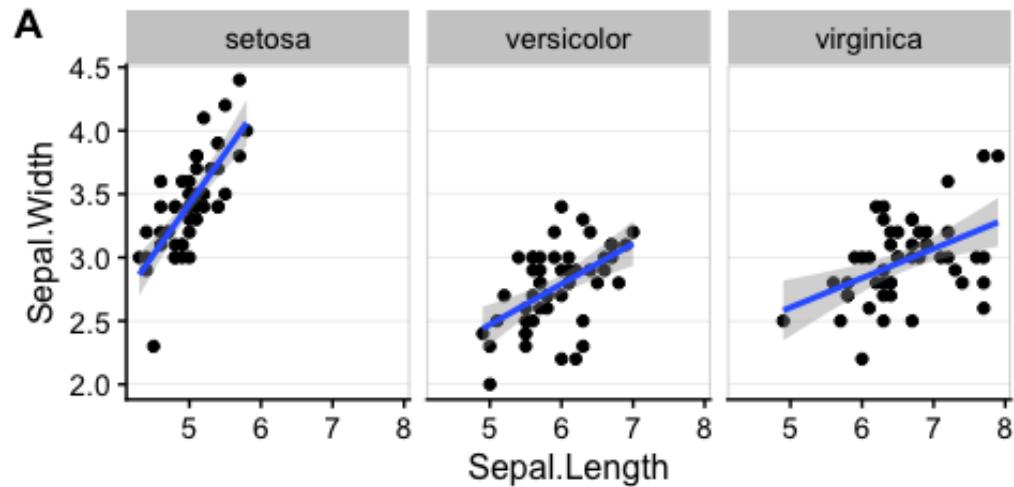
Script: This could be a range of file types, but generally a text file; your “analysis pipeline”



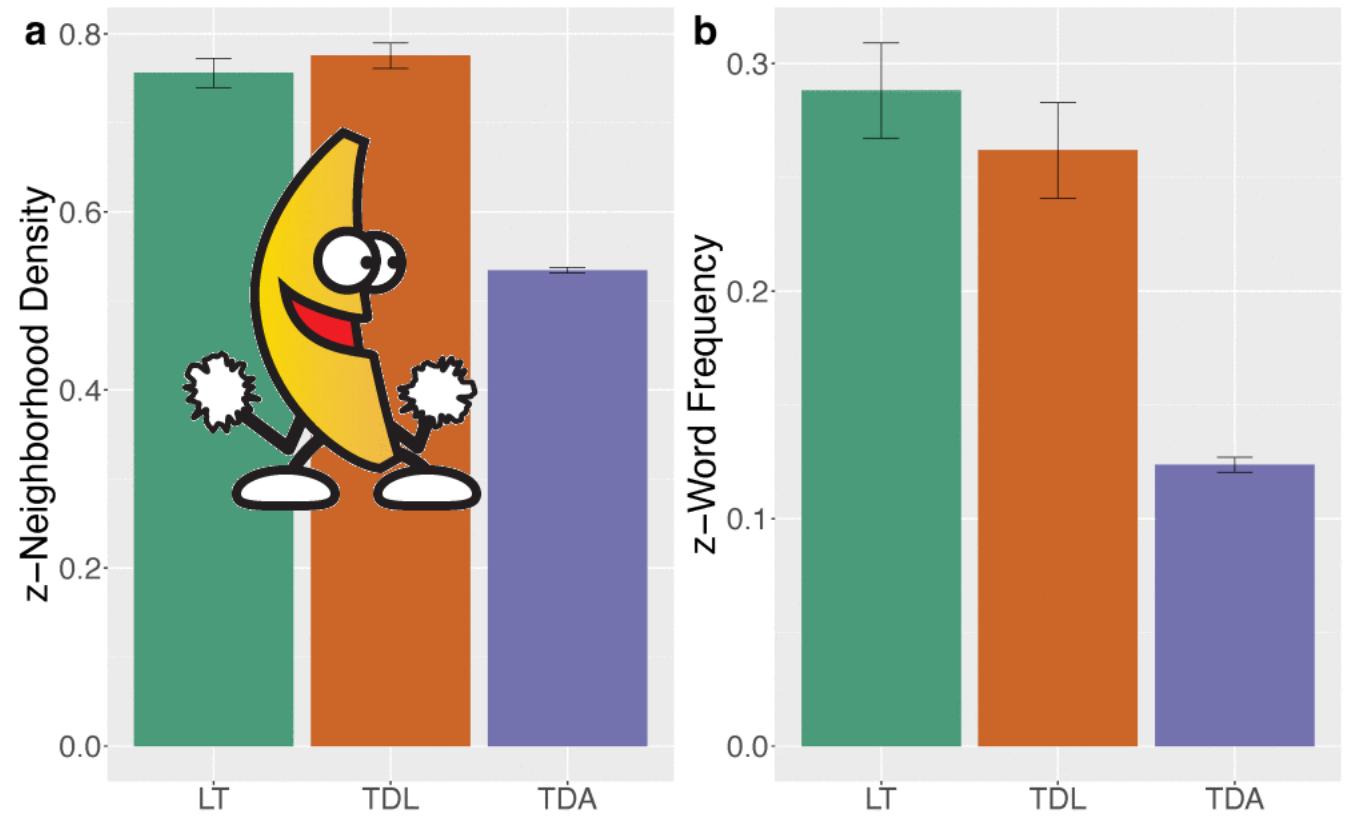
Packages

- Anything you can dream up – there's a package for that!
- Data manipulation (dplyr)
- Figures (ggplot, cowplot)
- Mixed effects models (lmer4)
- ANOVA (ez)
- Effect sizes (effsize)
- Open-source databases (childeadb)
- Image processing (magick)

ggplot2 (with cowplot)



magick



Advanced basics

- R is **case-sensitive** language
- Be consistent in your code – this allows you to recycle chunks of code and save time
 - Standardize coding conventions (i.e., subject number is always subject_id)
- Operations in R are performed on **data frames**
 - Think of them as a spreadsheet
 - Rows = observations
 - Columns/vectors = variables

data frame (df) called all.produces -
R doesn't like spaces!

all.produces data frame contains
563,901 observations and 35
variables

Corpus - RStudio

2023_02_13_CDI_Stats.Rmd | all | all.produces

Filter

data_id	age	sex	mom_ed	value	item_id	type	category	definition	UnTrn	StTrn
6	129242	27	Female	Some Secondary	produces	item_14	word	animals	animal	AE.N.AH.M.AH.L
7	129242	27	Female	Some Secondary	produces	item_16	word	animals	bear	B.EH.R
8	129242	27	Female	Some Secondary	produces	item_18	word	animals	bird	B.ER.D
9	129242	27	Female	Some Secondary	produces	item_19	word	animals	bug	B.AH.G
10	129242	27	Female	Some Secondary	produces	item_20	word	animals	bunny	B.AH.N.IY
11	129242	27	Female	Some Secondary	produces	item_22	word	animals	cat	K.AE.T
12	129242	27	Female	Some Secondary	produces	item_23	word	animals	chicken	CH.IH.K.AH.N
13	129242	27	Female	Some Secondary	produces	item_26	word	animals	dog	D.AO.G
14	129242	27	Female	Some Secondary	produces	item_28	word	animals	duck	D.AH.K
15	129242	27	Female	Some Secondary	produces	item_30	word	animals	fish	F.IH.SH
16	129242	27	Female	Some Secondary	produces	item_35	word	animals	horse	HH.AO.R.S
17	129242	27	Female	Some Secondary	produces	item_36	word	animals	kitty	K.IH.T.IY
18	129242	27	Female	Some Secondary	produces	item_39	word	animals	monkey	M.AH.NG.K.IY
19	129242	27	Female	Some Secondary	produces	item_41	word	animals	mouse	M.AW.S
20	129242	27	Female	Some Secondary	produces	item_43	word	animals	penguin	P.EH.NG.G.W.AH.N
21	129242	27	Female	Some Secondary	produces	item_45	word	animals	pony	P.OW.N.IY
22	129242	27	Female	Some Secondary	produces	item_46	word	animals	puppy	P.AH.P.IY
24	129242	27	Female	Some Secondary	produces	item_56	word	vehicles	airplane	EH.R.P.L.EY.N
25	129242	27	Female	Some Secondary	produces	item_57	word	vehicles	bicycle	B.AY.S.IH.K.AH.L
26	129242	27	Female	Some Secondary	produces	item_58	word	vehicles	boat	B.OW.T
27	129242	27	Female	Some Secondary	produces	item_59	word	vehicles	bus	B.AH.S

Environment | History | Connections | Tutorial | Import Dataset | 3.32 GB | Corpus

all.produces | 563901 obs. of 35 variables

```

$ data_id : Factor w/ 2439 levels "129242","129243",...: 1 1 1 1 1 1 1 1 1 ...
$ age : int 27 27 27 27 27 27 27 27 27 ...
$ sex : chr "Female" "Female" "Female" "Female" ...
$ mom_ed : chr "Some Secondary" "Some Secondary" "Some Secondary" "Some Secondary" ...
$ value : Factor w/ 7 levels "", "complex", "not yet", ...: 5 5 5 5 5 5 5 5 5 ...
$ item_id : chr "item_14" "item_16" "item_18" "item_19" ...
$ type : chr "word" "word" "word" "word" ...
$ category : chr "animals" "animals" "animals" "animals" ...
$ definition : Factor w/ 591 levels "about", "above", ...: 10 43 57 82 85 104 112 157 172 191 ...
$ UnTrn : chr "AE.N.AH.M.AH.L" "B.EH.R" "B.ER.D" "B.AH.G" ...
$ StTrn : chr "AE1.N.AH0.M.AH0.L" "B.EH1.R" "B.ER1.D" "B.AH1.G" ...
$ NPhon : int 6 3 3 3 4 3 5 3 3 3 ...
$ unsDENS : int 3 35 35 33 27 45 4 19 41 18 ...
$ unsFDEN : num 39.2 8934.1 1669.8 7305.8 3157.1 ...
$ unsLDEN : num 1.64 63.83 30.41 37.17 26.94 ...
$ unsCDEN : int 1141 71568 31767 36660 28964 55408 373 10057 31558 12809 ...
$ SFreq : num 45.5 57.4 45.5 20.9 18.6 ...
$ SCDcnt : int 1372 1500 1196 583 340 1386 1570 3049 703 1708 ...
$ Syll : Factor w/ 5 levels "1", "2", "3", "4", ...: 3 1 1 1 2 1 2 1 1 1 ...
$ gender : chr "female" "female" "female" "female" ...
$ vocab_nwords : int 497 497 497 497 497 497 497 497 ...
$ age_months : int 27 27 27 27 27 27 27 27 ...

```

Learn from my mistakes

- Practice writing code! Practice makes (almost) perfect!
- Don't reinvent the wheel – look for pre-existing code then adapt
- Re-analyze old data sets in R
- Good places to start:
 - Intro to data manipulation, visualization and basic analyses:
<http://www.cookbook-r.com/>
 - If you want a step-by-step intro:
<https://www.rstudio.com/resources/webinars/a-gentle-introduction-to-tidy-statistics-in-r/>
 - Got an error? <https://stackoverflow.com/> (watch out – responses can be snarky)

Next steps: Download R and R Studio

Step 1) Download R here: [r-project.org](https://www.r-project.org) & install

Step 2) Download RStudio here: posit.co & install