

# WRITING WITH DATA

Jeff Goldsmith, PhD  
Department of Biostatistics

# Writing is important

- You're going to spend a lot of your time communicating in writing
  - With collaborators, a general public, future you
  - About data cleaning, analyses, results
  - In formal reports, brief summaries, replies to questions
- Time to get good

# Tools

- Code is necessary but not sufficient
- Use tools that combine your code and text
- Greatly facilitates reproducibility, which is a big concept
  - In short, someone you don't know or work with should be able to reproduce each step of your analysis
  - As a part of this, they should understand why you did what you did
  - (Again, this someone is often future you)
- We'll use R Markdown to write reproducible reports

# General tips

- Know your audience
  - Are they statistically knowledgeable?
  - How many details do they want / need?
- Say exactly what you did
  - Don't leave any thing important out
  - Not the same as a step-by-step list of what you typed into R

# General structure

- Introduction / overview
- Data and methods
  - File names
  - Summary statistics
  - Exploratory analysis
  - Formal analysis
- Results
- Discussion
  
- Some version of these exist in almost everything I write
- Sometimes these are long, sometimes they're a sentence

# Introduction

- What is the context for this problem?
- What kind of data were gathered?
- What do you hope to learn?

# Data

- Importing, tidying, and editing
  - Loading data
  - Reorganizing into usable form
  - Identifying missing values
  - Recoding and creating variables
- Summary statistics
  - Sample size
  - Means or proportions of major variables

# Methods / “models”

- Exploratory analyses
  - Visualizations
  - Numerical summaries
- Formal analyses
  - Model components
  - Model strategy
  - Formal comparisons of interest, tests, significance levels



# Results

- What did you find in exploratory analyses (any missing values? data distributions? notable features?)
- What happened in your modeling?
- What is your final model, and what are the important quantities?

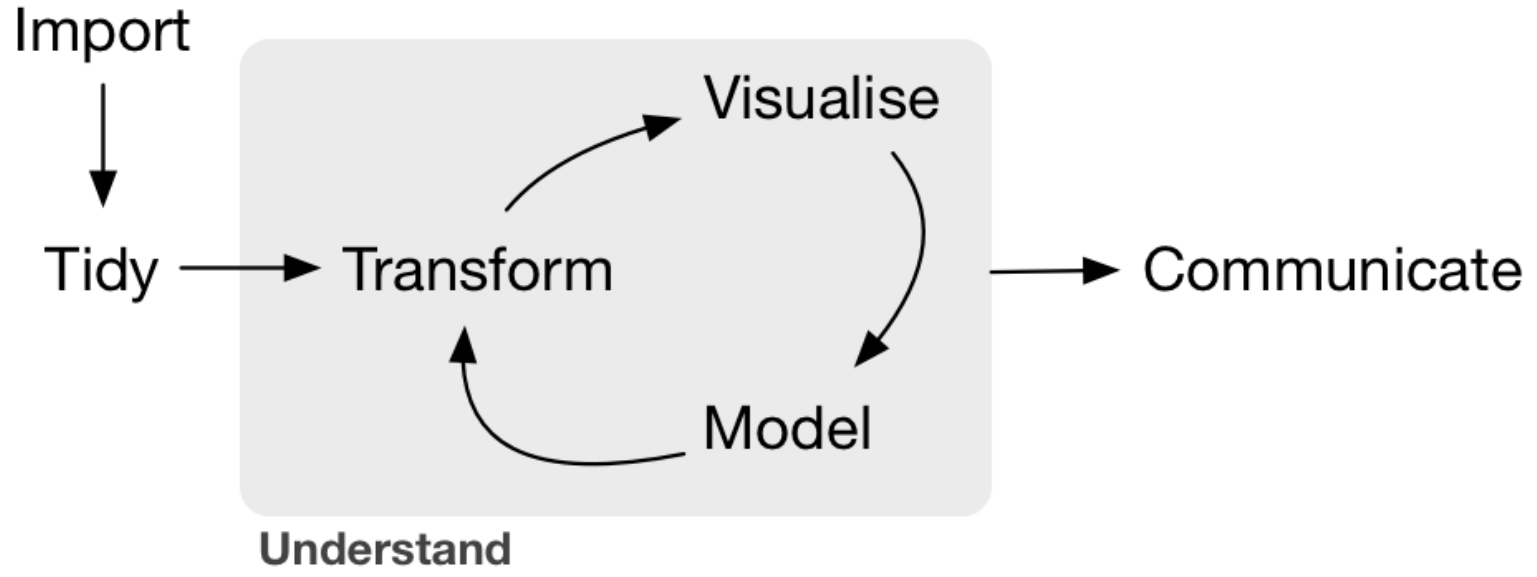
# Discussion

- What do your results say about the question you hoped to answer?
- What were the limitations of your data or your analysis?
- What open questions remain? Are any of these solvable with the current data?
- What are your next steps?

# Some true stuff about writing

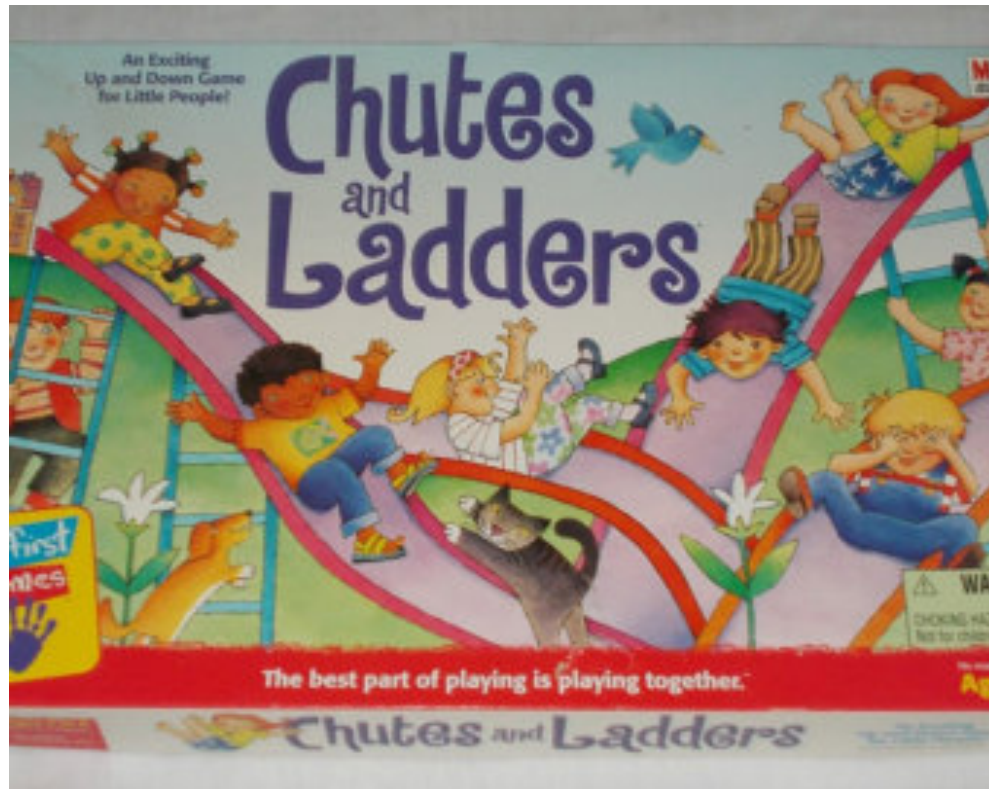
- It is not easy
- It takes practice
- It is critical to do well

# Recall ...

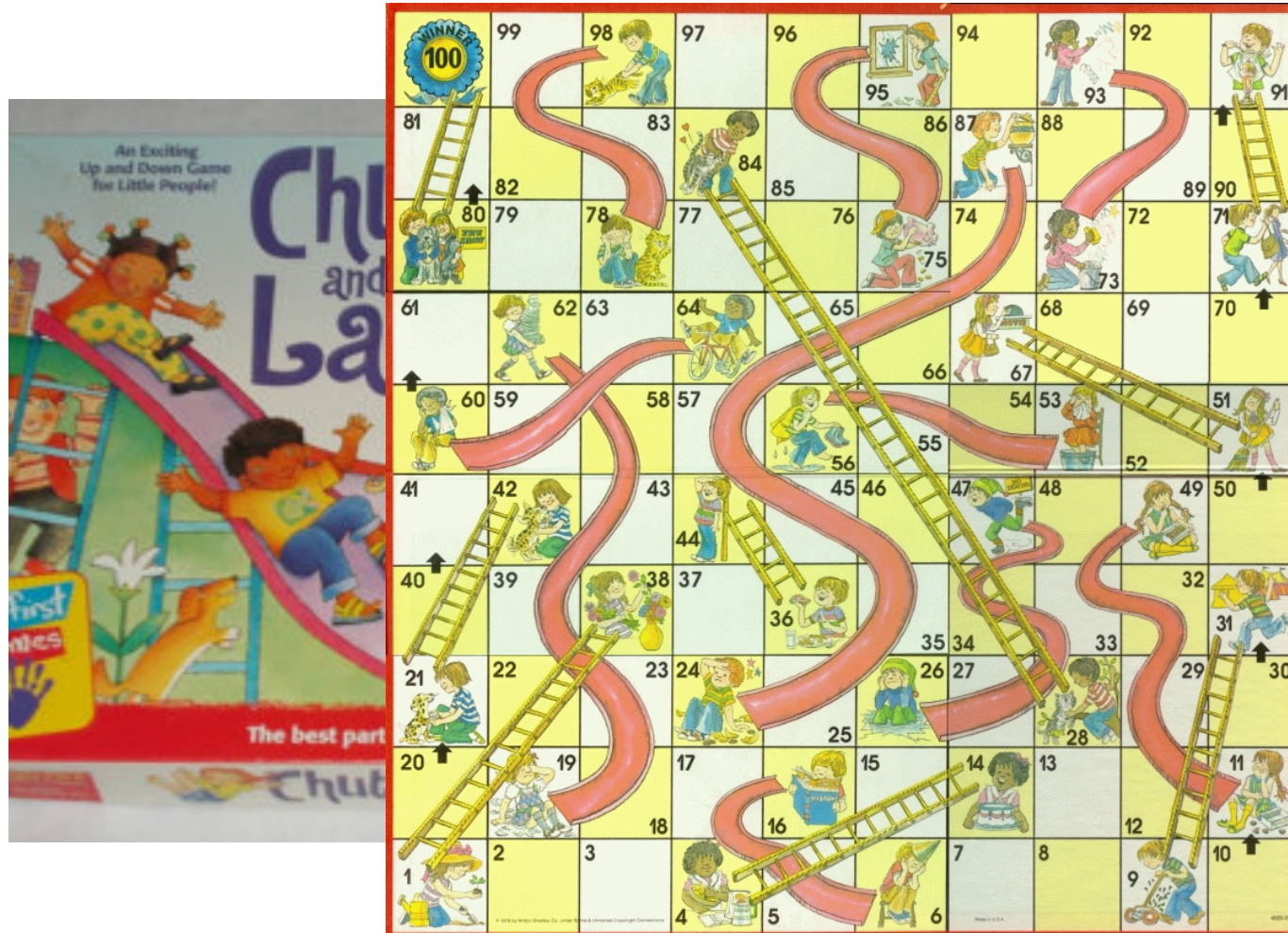


R for Data Science

# How analyses are in reality



# How analyses are in reality





# How analyses are presented



# Be complete ...

... but not too complete.

P811HW3

3/8/15, 2:50 PM

```
## [12,] -0.08394047 -0.032693143
## [13,] -0.08403121 -0.034394585
## [14,] -0.08397860 -0.034980378
## [15,] -0.08434735 -0.036156917
## [16,] -0.08435698 -0.036650398
## [17,] -0.08447201 -0.037679892
## [18,] -0.08412698 -0.037754627
## [19,] -0.08435342 -0.038675913
## [20,] -0.08403146 -0.038832864
## [21,] -0.08331837 -0.038555963
## [22,] -0.08311964 -0.038486428
## [23,] -0.08264245 -0.038160749
## [24,] -0.08228691 -0.037783216
## [25,] -0.08166935 -0.037351882
## [26,] -0.08118348 -0.036826247
## [27,] -0.08011525 -0.035707460
## [28,] -0.07967686 -0.035021461
## [29,] -0.07867025 -0.033853880
## [30,] -0.07820231 -0.033233691
## [31,] -0.07719687 -0.032100156
## [32,] -0.07682622 -0.031555885
## [33,] -0.07581413 -0.030493934
## [34,] -0.07555181 -0.029784794
## [35,] -0.07424381 -0.028273179
## [36,] -0.07362121 -0.027558268
## [37,] -0.07244494 -0.026111580
## [38,] -0.07187484 -0.025583699
## [39,] -0.07056949 -0.024287362
## [40,] -0.06983318 -0.023357617
## [41,] -0.06851434 -0.022093530
## [42,] -0.06782302 -0.021520307
## [43,] -0.06691565 -0.020101802
## [44,] -0.06638024 -0.019371473
## [45,] -0.06541473 -0.018089351
## [46,] -0.06504171 -0.017422153
## [47,] -0.06337542 -0.016354797
## [48,] -0.06284136 -0.015832022
## [49,] -0.06196227 -0.014421761
## [50,] -0.06160143 -0.013729093
## [51,] -0.06073027 -0.012330712
## [52,] -0.06035373 -0.011730336
## [53,] -0.05985272 -0.010589271
## [54,] -0.05976229 -0.010116416
## [55,] -0.05931431 -0.009188999
## [56,] -0.05919370 -0.008598205
## [57,] -0.05923929 -0.007870010
## [58,] -0.05884408 -0.007463907
```



# Be complete ...

... but not too complete.

PS111HW3	PS111HW3	3/6/15, 2:50 PM	3/6/15, 2:50 PM
## [12,] -0.08394047 -0.032693143	## [59,] -0.05828045 -0.006427910		
## [13,] -0.08403121 -0.034394585	## [60,] -0.05849033 -0.006043732		
## [14,] -0.08397860 -0.034980378	## [61,] -0.05889288 -0.005276025		
## [15,] -0.08434735 -0.036156917	## [62,] -0.05939320 -0.004944063		
## [16,] -0.08435698 -0.036650398	## [63,] -0.06050816 -0.004206706		
## [17,] -0.08447201 -0.037679892	## [64,] -0.06104112 -0.003886241		
## [18,] -0.08412698 -0.037754627	## [65,] -0.06174906 -0.003648505		
## [19,] -0.08435342 -0.038675913	## [66,] -0.06241750 -0.003800381		
## [20,] -0.08403146 -0.038832864	## [67,] -0.06358717 -0.003662780		
## [21,] -0.08331837 -0.038555963	## [68,] -0.06437141 -0.003635112		
## [22,] -0.08311964 -0.038486428	## [69,] -0.06606900 -0.003829907		
## [23,] -0.08264245 -0.038160749	## [70,] -0.06673738 -0.004057332		
## [24,] -0.08228691 -0.037783216	## [71,] -0.06808891 -0.004816457		
## [25,] -0.08166935 -0.037351882	## [72,] -0.06912238 -0.005185557		
## [26,] -0.08118348 -0.036826247	## [73,] -0.07133348 -0.005956885		
## [27,] -0.08011525 -0.035707460	## [74,] -0.07252225 -0.006506062		
## [28,] -0.07967686 -0.035021461	## [75,] -0.07551133 -0.007708559		
## [29,] -0.07867025 -0.033853880	## [76,] -0.07707811 -0.008382036		
## [30,] -0.07820231 -0.033233691	## [77,] -0.07985765 -0.009874331		
## [31,] -0.07719687 -0.032100156	## [78,] -0.08089292 -0.010706543		
## [32,] -0.07682622 -0.031555885	## [79,] -0.08369384 -0.012351292		
## [33,] -0.07581413 -0.030493934	## [80,] -0.08526338 -0.013185658		
## [34,] -0.07555181 -0.029784794	## [81,] -0.08823390 -0.014956617		
## [35,] -0.07424381 -0.028273179	## [82,] -0.08979044 -0.016127845		
## [36,] -0.07362121 -0.027558268	## [83,] -0.09304215 -0.018502398		
## [37,] -0.07244494 -0.026111580	## [84,] -0.09474251 -0.019530388		
## [38,] -0.07187484 -0.025583699	## [85,] -0.09829321 -0.022059960		
## [39,] -0.07056949 -0.024287362	## [86,] -0.10029588 -0.023170549		
## [40,] -0.06983318 -0.023357617	## [87,] -0.10406055 -0.025920217		
## [41,] -0.06851434 -0.022093530	## [88,] -0.10600856 -0.027101026		
## [42,] -0.06782302 -0.021520307	## [89,] -0.11050742 -0.030160113		
## [43,] -0.06691565 -0.020101802	## [90,] -0.11284952 -0.031782676		
## [44,] -0.06638024 -0.019371473	## [91,] -0.11767979 -0.035136968		
## [45,] -0.06541473 -0.018089351	## [92,] -0.12016788 -0.037081378		
## [46,] -0.06504171 -0.017422153	## [93,] -0.12527376 -0.040924296		
## [47,] -0.06337542 -0.016354797	## [94,] -0.12789947 -0.043164742		
## [48,] -0.06284136 -0.015832022	## [95,] -0.13313524 -0.047185703		
## [49,] -0.06196227 -0.014421761	## [96,] -0.13570119 -0.049292382		
## [50,] -0.06160143 -0.013729093	## [97,] -0.14100031 -0.053871825		
## [51,] -0.06073027 -0.012330712	## [98,] -0.14373346 -0.056295080		
## [52,] -0.06035373 -0.0111730336	## [99,] -0.14936673 -0.061355509		
## [53,] -0.05985272 -0.010589271	## [100,] -0.15226672 -0.063875523		
## [54,] -0.05976229 -0.010116416	## [101,] -0.15837583 -0.069434604		
## [55,] -0.05931431 -0.009188999	## [102,] -0.16170942 -0.072208042		
## [56,] -0.05919370 -0.008598205	## [103,] -0.16861004 -0.077481323		
## [57,] -0.05923929 -0.007870010	## [104,] -0.17215587 -0.080216911		
## [58,] -0.05884408 -0.007463907	## [105,] -0.17936529 -0.085934601		

# Be complete ...

... but not too complete.

P8111HW3 3/8/15, 2:50 PM


```
## [12,] -0.08394047 -0.032693143
## [13,] -0.08403121 -0.034394585
## [14,] -0.08397860 -0.034980378
## [15,] -0.08434735 -0.036156917
## [16,] -0.08435698 -0.036650398
## [17,] -0.08447201 -0.037679892
## [18,] -0.08412698 -0.037754627
## [19,] -0.08435342 -0.038675913
## [20,] -0.08403146 -0.038832864
## [21,] -0.08331837 -0.038555963
## [22,] -0.08311964 -0.038486428
## [23,] -0.08264245 -0.038160749
## [24,] -0.08228691 -0.037783216
## [25,] -0.08166935 -0.037351882
## [26,] -0.08118348 -0.036826247
## [27,] -0.08011525 -0.035707460
## [28,] -0.07967686 -0.035021461
## [29,] -0.07867025 -0.033853880
## [30,] -0.07820231 -0.033233691
## [31,] -0.07719687 -0.032100156
## [32,] -0.07682622 -0.031555885
## [33,] -0.07581413 -0.030493934
## [34,] -0.07555181 -0.029784794
## [35,] -0.07424381 -0.028273179
## [36,] -0.07362121 -0.027558268
## [37,] -0.07244494 -0.026111580
## [38,] -0.07187484 -0.025583699
## [39,] -0.07056949 -0.024287362
## [40,] -0.06983318 -0.023357617
## [41,] -0.06851434 -0.022093530
## [42,] -0.06782302 -0.021520307
## [43,] -0.06691565 -0.020101802
## [44,] -0.06638024 -0.019371473
## [45,] -0.06541473 -0.018089351
## [46,] -0.06504171 -0.017422153
## [47,] -0.06337542 -0.016354797
## [48,] -0.06284136 -0.015832022
## [49,] -0.06196227 -0.014421761
## [50,] -0.06160143 -0.013729093
## [51,] -0.06073027 -0.012330712
## [52,] -0.06035373 -0.011730336
## [53,] -0.05985272 -0.010589271
## [54,] -0.05976229 -0.010116416
## [55,] -0.05931431 -0.009188999
## [56,] -0.05919370 -0.008598205
## [57,] -0.05923929 -0.007870010
## [58,] -0.05884408 -0.007463907
```

P8111HW3 3/8/15, 2:50 PM

```
## [59,] -0.05828045 -0.006427910
## [60,] -0.05849033 -0.006043732
## [61,] -0.05889288 -0.005276025
## [62,] -0.05939320 -0.004944063
## [63,] -0.06050816 -0.004206706
## [64,] -0.06104112 -0.003886241
## [65,] -0.06174906 -0.003648505
## [66,] -0.06241750 -0.003800381
## [67,] -0.06358717 -0.003662780
## [68,] -0.06437141 -0.003635112
## [69,] -0.06606900 -0.003829907
## [70,] -0.06673738 -0.004057332
## [71,] -0.06808891 -0.004816457
## [72,] -0.06912238 -0.005185557
## [73,] -0.07133348 -0.005956885
## [74,] -0.07252225 -0.006506062
## [75,] -0.07551133 -0.007708559
## [76,] -0.07707811 -0.008382036
## [77,] -0.07985765 -0.009874331
## [78,] -0.08089292 -0.010706543
## [79,] -0.08369384 -0.012351292
## [80,] -0.08526338 -0.013185658
## [81,] -0.08823390 -0.014956617
## [82,] -0.08979044 -0.016127845
## [83,] -0.09304215 -0.018502398
## [84,] -0.09474251 -0.019530388
## [85,] -0.09829321 -0.022059960
## [86,] -0.10029588 -0.023170549
## [87,] -0.10406055 -0.025920217
## [88,] -0.10600856 -0.027101026
## [89,] -0.11050742 -0.030160113
## [90,] -0.11284952 -0.031782676
## [91,] -0.11767979 -0.035136968
## [92,] -0.12016788 -0.037081378
## [93,] -0.12527376 -0.040924296
## [94,] -0.12789947 -0.043164742
## [95,] -0.13313524 -0.047185703
## [96,] -0.13570119 -0.049292382
## [97,] -0.14100031 -0.053871825
## [98,] -0.14373346 -0.056295080
## [99,] -0.14936673 -0.061355509
## [100,] -0.15226672 -0.063875523
## [101,] -0.15837583 -0.069434604
## [102,] -0.16170942 -0.072208042
## [103,] -0.16861004 -0.077481323
## [104,] -0.17215587 -0.080216911
## [105,] -0.17936529 -0.085934601
```

P8111HW3 3/8/15, 2:50 PM

```
## [200,] -0.78579562 -0.671141895
## [201,] -0.78649970 -0.670452918
## [202,] -0.78675680 -0.669902358
## [203,] -0.78512608 -0.668247942
## [204,] -0.78453474 -0.666857604
## [205,] -0.78443188 -0.662315516
## [206,] -0.78302324 -0.661049715
## [207,] -0.78035282 -0.657166622
## [208,] -0.77942284 -0.654408116
## [209,] -0.77706316 -0.647297818
## [210,] -0.77494542 -0.644028816
## [211,] -0.77142512 -0.637076968
## [212,] -0.76967333 -0.632319394
## [213,] -0.76557589 -0.623728406
## [214,] -0.76276779 -0.619648160
## [215,] -0.75721274 -0.609000925
## [216,] -0.75211326 -0.603117713
## [217,] -0.74700729 -0.589563950
## [218,] -0.74318011 -0.583664200
## [219,] -0.73463678 -0.567231414
## [220,] -0.73067486 -0.559006529
## [221,] -0.72235137 -0.541947505
```



Confidence intervals and the fitted value curve in each bootstrap sample is done above.

Part e)

```
### plot them using ggplot
eachcil<-as.data.frame(eachci)
plot = ggplot() + geom_point(color = "black") +
  geom_line(aes(x=lidar$range, y = UB), color = "blue", linetype = 2) +
  geom_line(aes(x=lidar$range, y = LB), color = "blue", linetype = 2) +
  geom_line(data=eachcil, aes(x=lidar$range, y = V2), color = "green", linetype = 2) +
  geom_line(data=eachcil, aes(x=lidar$range, y = V1), color = "green", linetype = 2)
plot
```

# Striking a balance

- This is where practice comes in

# R Markdown?

- A “Markdown” language is a lightweight syntax that can be easily converted to HTML or another format (PDF, Word)
- R Markdown lets you combine formatted text with code chunks and the results of those chunks



- Having text and code in the same place, and having the combined output be user-friendly, is huge for your workflow

# R Markdown?

- A “Markdown” language is a lightweight syntax that can be easily converted to HTML or another format (PDF, Word)
- R Markdown lets you combine formatted text with code chunks and the results of those chunks



- Having text and code in the same place, and having the combined output be user-friendly, is huge for your workflow