

Bush 631-600: Quantitative Methods

Lecture 1 (08.30.2022): Introduction

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Fall 2022

What is today's plan?

- ▶ Introductions.
- ▶ What is this class?
- ▶ Syllabus 'deep dive'.
- ▶ Programming with R.

Introductions

- ▶ Rotem Dvir: PhD in political science (Texas A&M, 2021).
 - ▶ Major field: International Relations.
 - ▶ Focus: International security and foreign policy.
-
- ▶ Assistant research scientist - ISTPP, Bush School.
 - ▶ Study public policy: health care, tech, critical infrastructure.
 - ▶ More? Check my website ([Link](#))

Bush 631: What are we doing here?

- ▶ Quantitative social science.
- ▶ Investigate social, economic and political world.

The collage illustrates the application of quantitative social science across various fields:

- Top Left:** A computer screen displaying a snippet of JavaScript code, specifically a function for handling form inputs. The code uses document.getElementById and var statements to manipulate the DOM.
- Top Right:** A soldier in full combat gear stands next to a tan military Humvee in an urban environment. This represents the use of quantitative methods in military and security studies.
- Bottom Left:** Seven international leaders standing in front of their respective flags (China, France, Germany, European Union, Iran, Russia, and the United States). This image represents the use of quantitative methods in international relations and comparative politics.
- Bottom Right:** A collection of data visualization dashboards. The top section shows a chart titled "GDP Growth by Country" with a blue line graph and a bar chart. Below are three main sections: "Who donates to whom and for what purpose?", "Donations Around the World: Summary", and "All donations to India within Asia Pacific". Each section contains various charts, maps, and data tables.

Quantitative Social Science

- ▶ Data science: Facebook, Twitter, Tiktok, Google
- ▶ Non-profits, government agencies: conduct policy evaluation with data.



Researcher

National Football League (NFL) · Culver City, CA · 2 weeks ago · 33 applicants



Full-time · Associate



Research Analyst - Advertiser Measurement

Spotify · New York, NY · 6 days ago · [21 applicants](#)



Full-time

Social Science Research Analyst

US Administration for Children and Families
Washington, DC
\$103,690 - \$134,798 a year · Full-time

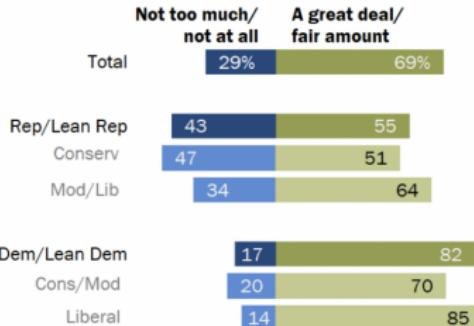
Data Analyst, Government

Civis Analytics 5 reviews
Illinois · Remote

Studying global issues with data

Democrats more likely to believe the U.S. benefits from NATO membership

% who think the U.S. benefits ___ from being a member of NATO



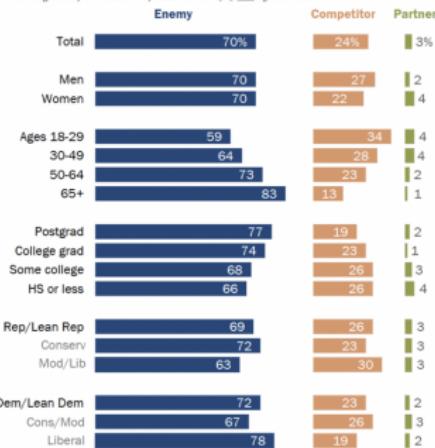
Note: Those who did not answer not shown.

Source: Survey of U.S. adults conducted March 21-27, 2022. Q62a.
“Seven-in-Ten Americans Now See Russia as an Enemy”

PEW RESEARCH CENTER

Older and more educated Americans more likely to see Russia as an enemy

% who say that, on balance, Russia is a(n) ___ of the U.S.



Note: Those who did not answer not shown.

Source: Survey of U.S. adults conducted March 21-27, 2022. Q61.
“Seven-in-Ten Americans Now See Russia as an Enemy”

PEW RESEARCH CENTER

Data driven Policymaking

Human choices research
and data

Status-quo bias

Default options

Save for retirement:
opt-out or opt-in?

Richard H. Thaler
Cass R. Sunstein

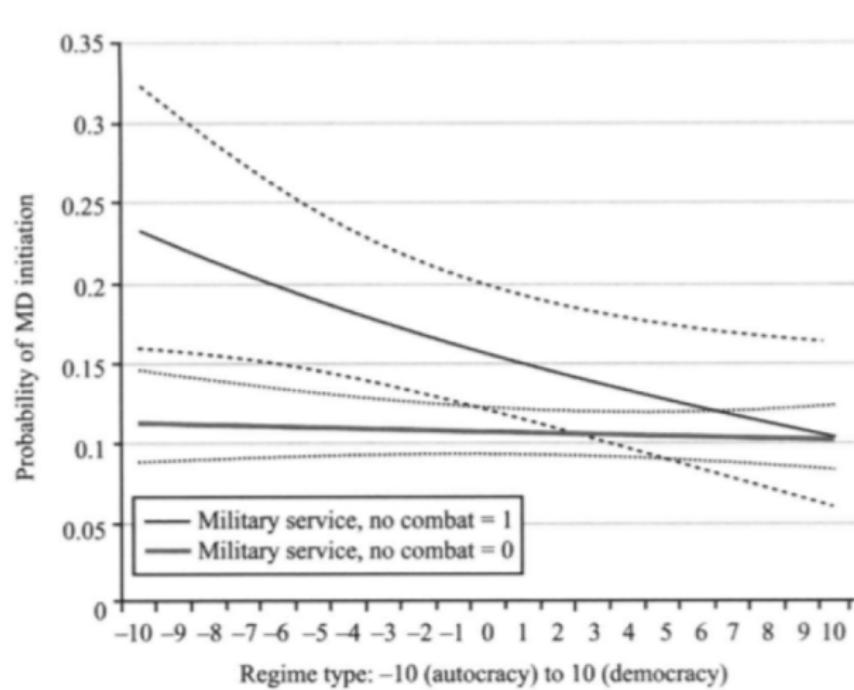
Nudge



Improving Decisions
About Health, Wealth,
and Happiness

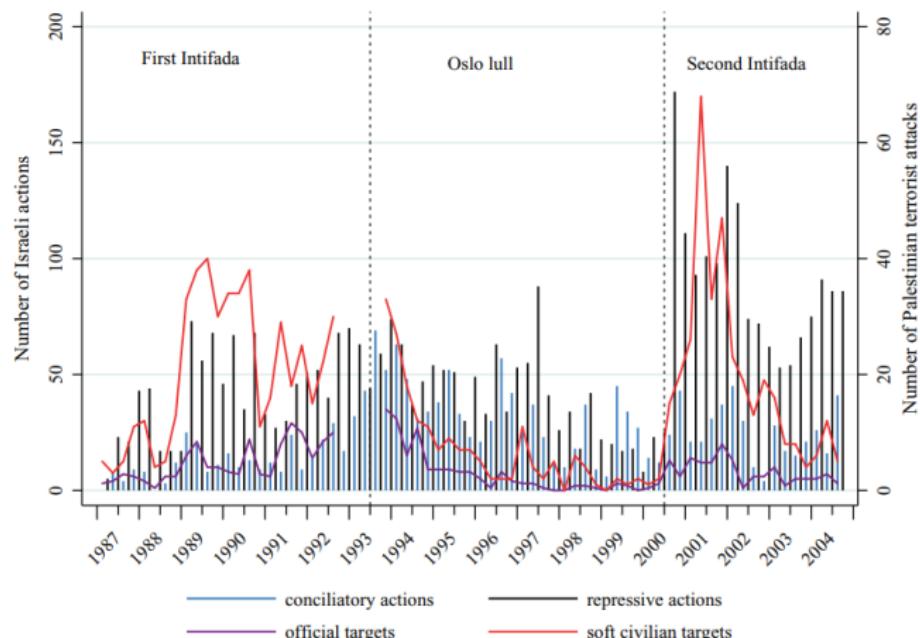
International Affairs research

- ▶ What drives the likelihood of international conflicts?
- ▶ The leaders? or the regimes?



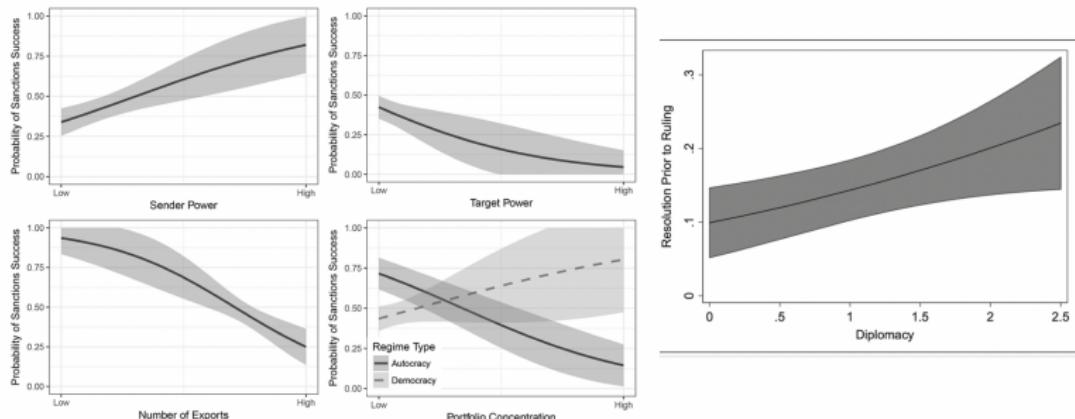
International Affairs research

- ▶ Counter terrorism strategies and organizational responses.



International Affairs research

- ▶ Diplomacy; Political economy: global trade flows, sanctions, foreign aid.



Quant methods

STUDY PUBLIC POLICY WITH STATS AND MATH...



How many of
you are here?

And here? →

$$\frac{\partial}{\partial \theta} \ln f_{a,\sigma^2}(\xi_1) = \frac{(\xi_1 - a)}{\sigma^2} f_{a,\sigma^2}(\xi_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\xi_1-a)^2}{2\sigma^2}}$$
$$\int T(x) \cdot \frac{\partial}{\partial \theta} f(x, \theta) dx = M \left(T(\xi), \frac{\partial}{\partial \theta} \ln L(\xi, \theta) \right)$$
$$\int T(x) \cdot \left(\frac{\partial}{\partial \theta} \ln L(x, \theta) \right) \cdot f(x, \theta) dx = \int T(\xi) \left(\frac{\frac{\partial}{\partial \theta} f(\xi, \theta)}{f(\xi, \theta)} \right) d\xi$$
$$\frac{\partial}{\partial \theta} \ln T(\xi) = \frac{\partial}{\partial \theta} \int T(x) f(x, \theta) dx = \int \frac{\partial}{\partial \theta} T(x) f(x, \theta) dx$$
A black silhouette of a person running away from a large, complex mathematical equation that is floating in the air. The equation is identical to the one above it, showing the derivative of the log-likelihood function with respect to theta.

Quant methods

Why? Why would I do that?



Why should I take this class?

- ▶ Skills to explore important questions:
 - ▶ Do economic sanctions work? when? why not?
 - ▶ Who support and who rejects free trade? why?
 - ▶ Why does leader decapitation limited as a counter-terrorism strategy?
- ▶ Transferable skills across industries:
 - ▶ Programming.
 - ▶ Data analysis.
 - ▶ Design social science research.
 - ▶ Writing professional docx: reports, briefs, executive summary.
 - ▶ Visualizing - plots, figures, infographics.
- ▶ You have to? :)

How are we doing it?

- ▶ Two interconnected paths:
 1. Theory and research design.
 2. Hands-on data analysis using R.

Important to remember

First - we design our research to answer the question(s)...

Then, we use programming (with R) for analysis

We end-up with (hopefully...) relevant insights

Syllabus 'deep dive'

Remember - it's all in the syllabus . . .



Syllabus ‘deep dive’

The essentials:

- ▶ When: Tuesdays, 4:30-7:20pm
- ▶ Where: Allen 1003

Office hours:

- ▶ Tuesday & Thursday 9:30-11:00am.
- ▶ My office: Allen 3029.
- ▶ Email, Zoom meeting.

Why office hours? Have you watched? ([Link](#))

Syllabus ‘deep dive’

Lectures:

- ▶ Be ready - read before class.
- ▶ Required readings - QSS book (all marked on syllabus).
- ▶ Recommended readings - should I read? what are they?
- ▶ Taking notes in class.
- ▶ Programming practice: HW and class assignments.

Course material and resources:

- ▶ Website: slides, R code, tasks instructions, other resources.
- ▶ Canvas: announcements, assignments, course material.
- ▶ Email if you need help.

Syllabus ‘deep dive’

Attendance:

- ▶ Joint effort to learn methods.
- ▶ 5% of final grade.
- ▶ 1 unexcused absence. Then what?
- ▶ Notify me **before** you miss class.

Remember COVID???

- ▶ Please don't come to class if you're sick/tested positive.
- ▶ Follow A&M guidelines (forms, quarantine).
- ▶ University accepted excused absence.

Syllabus ‘deep dive’

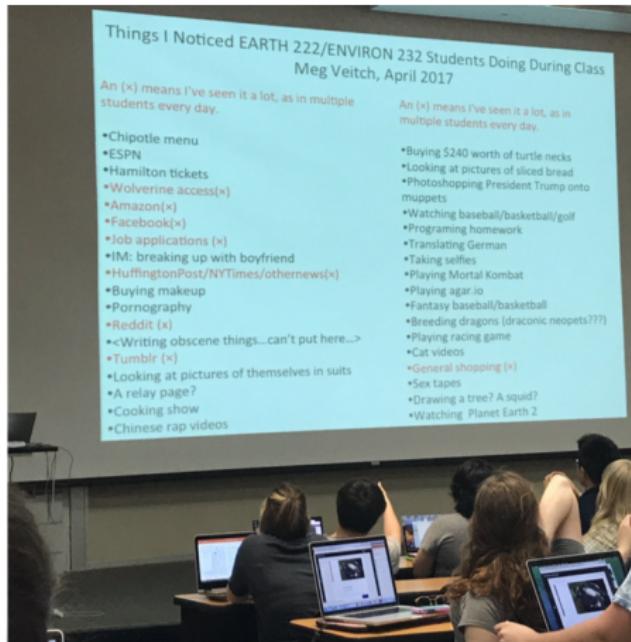
Grading and assignments:

1. Attendance (5%).
2. Home assignments (Swirl - R) (10%).
3. Research design task with R (10%).
4. Research design in class - 4 tasks (10%).
5. Course project - proposal (10%).
6. Course project - data report (15%).
7. Course project - poster/infographic (25%)
8. Course project - executive summary (10%).
9. Course project - peer review and feedback (5%).

Syllabus ‘deep dive’

Other issues:

- ▶ Make-up policy; Plagiarism.
- ▶ Electronics in class.



Syllabus ‘deep dive’

Topics overview:

1. Introduction and R basics.
2. Causality (2 weeks).
3. Measurement (2 weeks).
4. Prediction (3 weeks).
5. Probability (2 weeks).
6. Estimation and uncertainty (3 weeks).
7. Summary.

Introduction to R

- ▶ Why R, isn't excel easier?
- ▶ Well...

Me: 12

Excel: 12

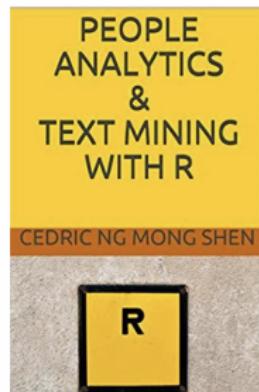
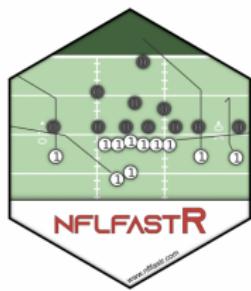
5

5

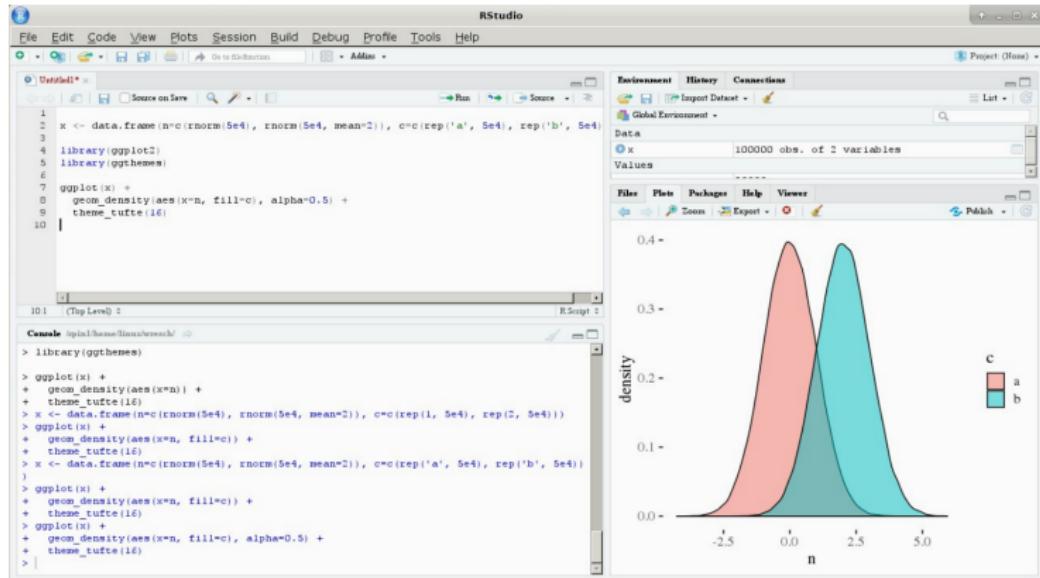
1/12/1900 12:00:00 PM

City	State	Abbreviation
Austin	TX	Texas
Salt Lake City	UT	Utah
Durham	NC	North Carolina
Columbus	OH	Orth Carolina
Baton Rouge	LA	Lorth Carolina
Omaha	NE	North Carolina
New Orleans	LA	Lorth Carolina
Des Moines	IA	Iorth Carolina
Seattle	WA	Worth Carolina
Oklahoma City	OK	Orth Carolina
Houston	TX	Torth Carolina
Charleston	SC	Sorth Carolina
Washington	DC	Dorth Carolina
Milwaukee	WI	Worth Carolina
Columbia	SC	Sorth Carolina
San Diego	CA	Corth Carolina
Orlando	FL	Forth Carolina
Boston	MA	Morth Carolina
Dallas	TX	Torth Carolina
Minneapolis	MN	Morth Carolina

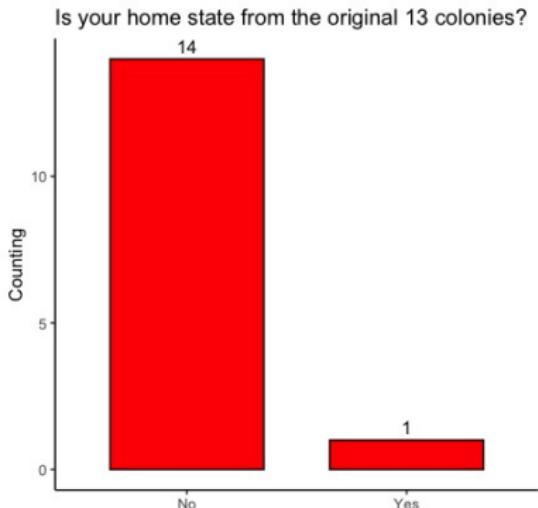
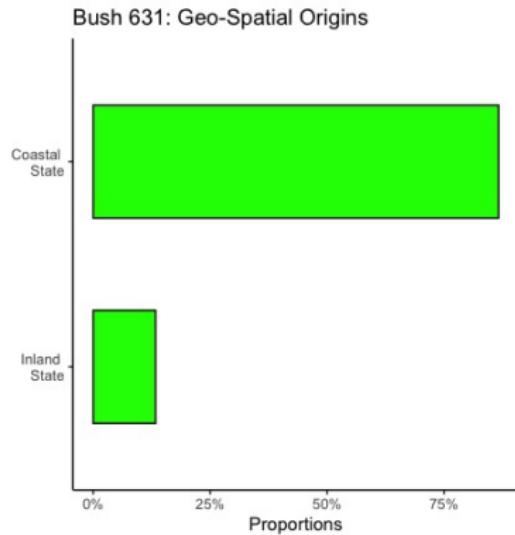
Introduction to R



Introduction to R

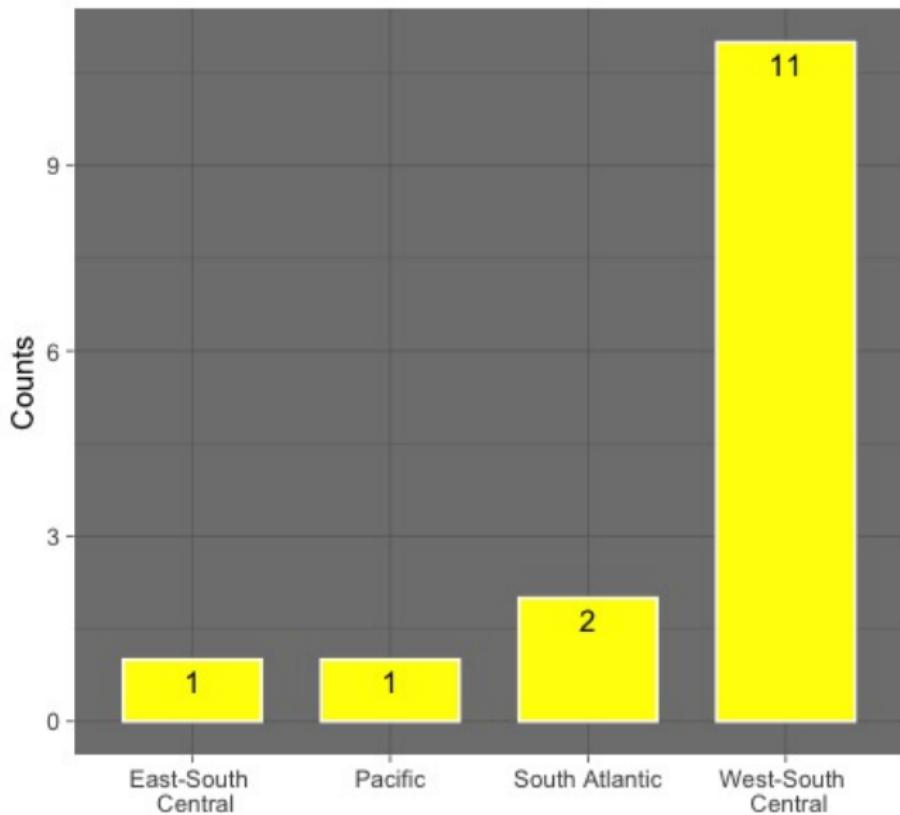


Syllabus Task with R

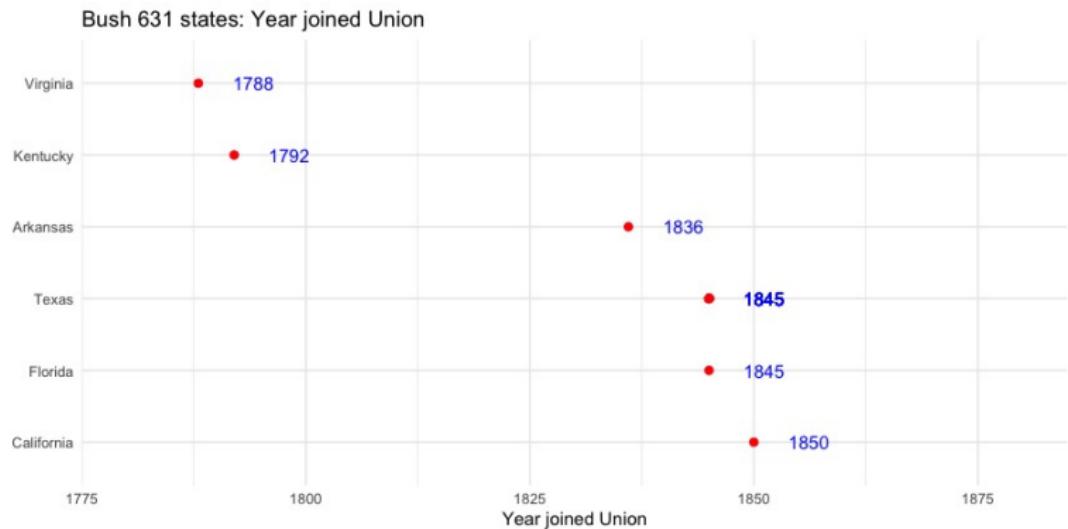


Syllabus Task with R

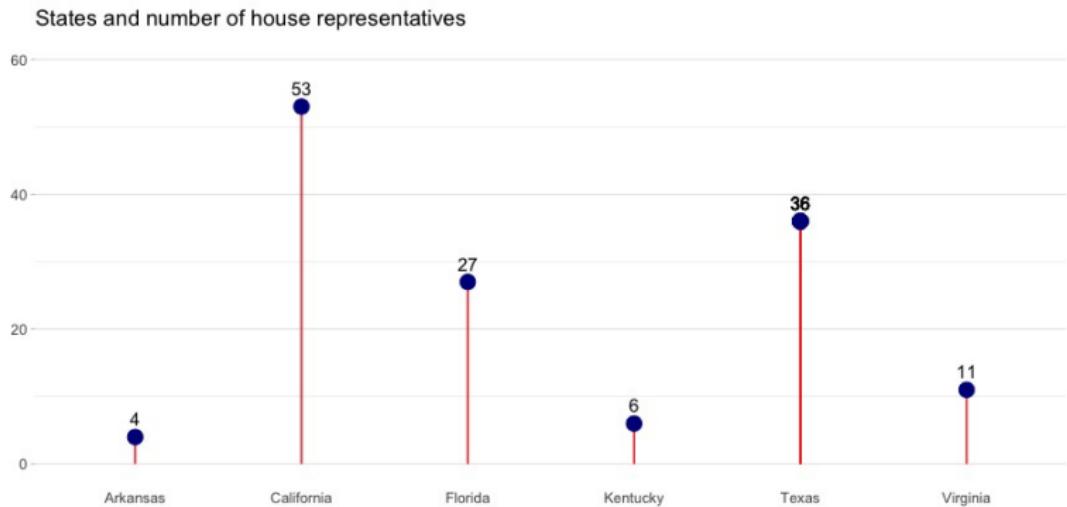
Your home state: Census 9 Divisions



Syllabus Task with R

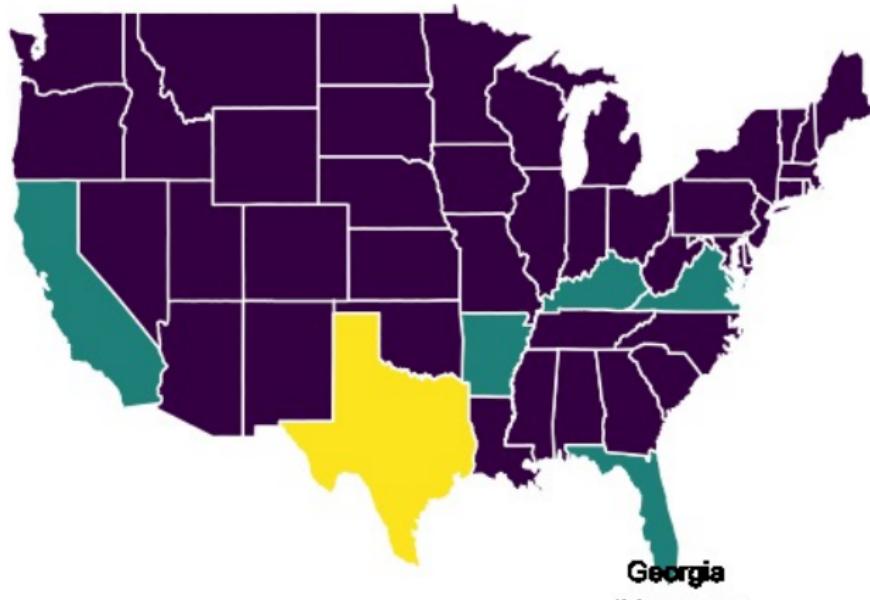


Syllabus Task with R



Syllabus Task with R

Bush 631: mapping home states



Class members in US states █ 0 █ 1 █ 10

Programming with R

Basic math tool

```
255+345
```

```
## [1] 600
```

```
255*345
```

```
## [1] 87975
```

```
255/345
```

```
## [1] 0.7391304
```

```
345/(2*255)
```

```
## [1] 0.6764706
```

```
sqrt(255)
```

```
## [1] 15.96872
```

Programming with R

Objects: storing information (number, string)

```
number <- 5
```

```
number
```

```
## [1] 5
```

```
no_number <- "5"
```

```
no_number
```

```
## [1] "5"
```

```
letter <- "W"
```

```
letter
```

```
## [1] "W"
```

```
word <- "Aggies"
```

```
word
```

```
## [1] "Aggies"
```

Programming with R

Objects: math results

```
result <- 2+5
```

```
result
```

```
## [1] 7
```

```
sqrt(result)
```

```
## [1] 2.645751
```

Potential Errors

```
no_number/5
```

Error in no_number/5 : non-numeric argument to binary operator

```
Result/5
```

Error: object 'Result' not found

Class: category/type of object

```
class(result)
## [1] "numeric"

class(letter)
## [1] "character"

class(word)
## [1] "character"

class(sqrt)
## [1] "function"
```

Vectors: array to store data

```
v1 <- c(1,2,3,4)
```

```
v1
```

```
## [1] 1 2 3 4
```

```
v2 <- c("A","B","C","D")
```

```
v2
```

```
## [1] "A" "B" "C" "D"
```

```
v3 <- c(11,12,13)
```

```
v_join1 <- c(v1,v2)
```

```
v_join1
```

```
## [1] "1" "2" "3" "4" "A" "B" "C" "D"
```

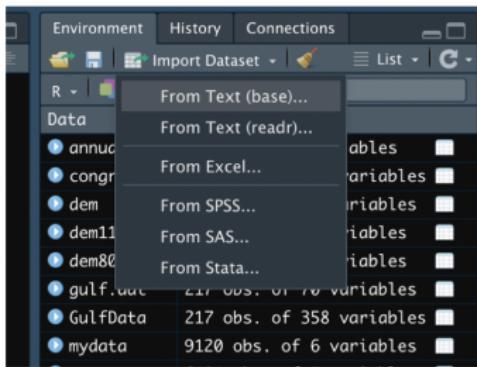
```
v_join2 <- c(v1,v3)
```

```
v_join2
```

```
## [1] 1 2 3 4 11 12 13
```

Working with data

1) Drop-down menu



2) Direct coding

```
190 library(readxl)
191 ags <- read_excel("~/Dropbox/TAMU/Bush631_QuantMethods/ags.xlsx")
192 View(ags)
193
```

- (a) Use the `read_` function
- (b) `View` – opens the data file for viewing

Our data

	Year	Wins	Losses	Pct	Coaching	Bowl
1	2021	8	4	0.667	Jimbo Fisher (8-4)	Gator Bowl-Did not play
2	2020	9	1	0.900	Jimbo Fisher (9-1)	Orange Bowl-W
3	2019	8	5	0.615	Jimbo Fisher (8-5)	Texas Bowl-W
4	2018	9	4	0.692	Jimbo Fisher (9-4)	Gator Bowl-W
5	2017	7	6	0.538	Jeff Banks (0-1), Kevin Sumlin (7-5)	Belk Bowl-L
6	2016	8	5	0.615	Kevin Sumlin (8-5)	Texas Bowl-L
7	2015	8	5	0.615	Kevin Sumlin (8-5)	Music City Bowl-L
8	2014	8	5	0.615	Kevin Sumlin (8-5)	Liberty Bowl-W
9	2013	9	4	0.692	Kevin Sumlin (9-4)	Chick-fil-A Bowl-W
10	2012	11	2	0.846	Kevin Sumlin (11-2)	Cotton Bowl-W
11	2011	7	6	0.538	Mike Sherman (6-6), Tim DeRuyter (1-0)	Meineke Car Care Bowl of Texas-W
12	2010	9	4	0.692	Mike Sherman (9-4)	Cotton Bowl-L
13	2009	6	7	0.462	Mike Sherman (6-7)	Independence Bowl-L
14	2008	4	8	0.333	Mike Sherman (4-8)	N/A
15	2007	7	6	0.538	Dennis Franchione (7-6)	Alamo Bowl-L
16	2006	9	4	0.692	Dennis Franchione (9-4)	Holiday Bowl-L
17	2005	5	6	0.455	Dennis Franchione (5-6)	N/A
18	2004	7	5	0.583	Dennis Franchione (7-5)	Cotton Bowl-L
19	2003	4	8	0.333	Dennis Franchione (4-8)	N/A
20	2002	6	6	0.500	R.C. Slocum (6-6)	N/A

Working with data: Indexing

```
ags[1,]

## # A tibble: 1 x 6
##   Year Wins Losses Pct Coaching      Bowl
##   <dbl> <dbl> <dbl> <dbl> <chr>
## 1 2021     8     4 0.667 Jimbo Fisher (8-4) Gator Bowl-Did not play
ags[,1]

## # A tibble: 119 x 1
##   Year
##   <dbl>
## 1 2021
## 2 2020
## 3 2019
## 4 2018
## 5 2017
## 6 2016
## 7 2015
## 8 2014
## 9 2013
## 10 2012
## # ... with 109 more rows
## # i Use `print(n = ...)` to see more rows
```

Working with data: Indexing

```
ags[c(1,2,4),]

## # A tibble: 3 x 6
##   Year   Wins Losses   Pct Coaching      Bowl
##   <dbl> <dbl> <dbl> <dbl> <chr>        <chr>
## 1 2021     8     4 0.667 Jimbo Fisher (8-4) Gator Bowl-Did not play
## 2 2020     9     1 0.9   Jimbo Fisher (9-1) Orange Bowl-W
## 3 2018     9     4 0.692 Jimbo Fisher (9-4) Gator Bowl-W
ags[1:3]

## # A tibble: 119 x 3
##   Year   Wins Losses
##   <dbl> <dbl> <dbl>
## 1 2021     8     4
## 2 2020     9     1
## 3 2019     8     5
## 4 2018     9     4
## 5 2017     7     6
## 6 2016     8     5
## 7 2015     8     5
## 8 2014     8     5
## 9 2013     9     4
## 10 2012    11    2
## # ... with 109 more rows
## # i Use `print(n = ...)` to see more rows
```

Working with data: Indexing

```
ags[c("Coaching")]
```

```
## # A tibble: 119 x 1
##   Coaching
##   <chr>
## 1 Jimbo Fisher (8-4)
## 2 Jimbo Fisher (9-1)
## 3 Jimbo Fisher (8-5)
## 4 Jimbo Fisher (9-4)
## 5 Jeff Banks (0-1), Kevin Sumlin (7-5)
## 6 Kevin Sumlin (8-5)
## 7 Kevin Sumlin (8-5)
## 8 Kevin Sumlin (8-5)
## 9 Kevin Sumlin (9-4)
## 10 Kevin Sumlin (11-2)
## # ... with 109 more rows
## # i Use `print(n = ...)` to see more rows
ags[1:10, c("Year", "Bowl")]
```

```
## # A tibble: 10 x 2
##   Year Bowl
##   <dbl> <chr>
## 1 2021 Gator Bowl-Did not play
## 2 2020 Orange Bowl-W
## 3 2019 Texas Bowl-W
## 4 2018 Gator Bowl-W
## 5 2017 Belk Bowl-L
## 6 2016 Texas Bowl-L
## 7 2015 Music City Bowl-L
## 8 2014 Liberty Bowl-W
## 9 2013 Chick-fil-A Bowl-W
## 10 2012 Cotton Bowl-W
```

Working with data: *using the \$ sign*

```
ags$Year[5]  
  
## [1] 2017  
ags$Coaching[1:5]  
  
## [1] "Jimbo Fisher (8-4)"  
## [2] "Jimbo Fisher (9-1)"  
## [3] "Jimbo Fisher (8-5)"  
## [4] "Jimbo Fisher (9-4)"  
## [5] "Jeff Banks (0-1), Kevin Sumlin (7-5)"
```

Math operations and data vectors

```
ags_win_p <- ags$Pct * 100
```

```
ags_win_p
```

```
## [1] 66.7 90.0 61.5 69.2 53.8 61.5 61.5 61.5 69.2 84.6 53.8 69.2
## [13] 46.2 33.3 53.8 69.2 45.5 58.3 33.3 50.0 66.7 58.3 66.7 78.6
## [25] 69.2 50.0 75.0 95.5 83.3 92.3 83.3 73.1 66.7 58.3 83.3 75.0
## [37] 83.3 54.5 50.0 45.5 58.3 36.4 54.5 66.7 66.7 83.3 83.3 72.7
## [49] 45.5 27.3 45.5 18.2 30.0 30.0 63.6 45.0 30.0 10.0 25.0 30.0
## [61] 45.0 25.0 30.0 40.0 72.7 95.0 75.0 10.0 45.0 35.0 60.0 63.6
## [73] 15.0 5.0 35.0 40.0 60.0 63.6 75.0 45.0 81.8 90.0 100.0 50.0
## [85] 66.7 70.8 30.0 27.3 65.0 50.0 70.0 22.2 55.6 55.0 94.4 61.1
## [97] 83.3 75.0 61.1 55.6 77.8 81.3 100.0 85.7 100.0 66.7 75.0 81.3
## [109] 44.4 88.9 85.7 88.9 93.8 37.5 81.3 85.7 77.8 66.7 68.2
```

Functions

Multiple functions for data summary:

- ▶ length (of vector)
- ▶ min & max values (for the whole vector)
- ▶ mean
- ▶ range
- ▶ sum

Functions: code examples

```
length(ags)

## [1] 6
min(ags$Losses)

## [1] 0
max(ags$Wins)

## [1] 12
mean(ags$Wins)

## [1] 6.252101
mean(ags$Pct)

## [1] 0.6058151
range(ags$Wins)

## [1] 0 12
range(ags$Coaching)

## [1] "Bear Bryant (1-9)"      "Walter Bachman (7-2)"
sum(ags$Wins) / length(ags$Wins)

## [1] 6.252101
```

Functions: code examples

```
sec <- seq(from = 2012, to = 2021, by = 1)
sec_coach <- ags$Coaching[10:1]
names(sec_coach) <- sec
sec_coach

##                               2012                               2013
## "Kevin Sumlin (11-2)"      "Kevin Sumlin (9-4)"
##                               2014                               2015
## "Kevin Sumlin (8-5)"       "Kevin Sumlin (8-5)"
##                               2016                               2017
## "Kevin Sumlin (8-5)" "Jeff Banks (0-1), Kevin Sumlin (7-5)"
##                               2018                               2019
## "Jimbo Fisher (9-4)"      "Jimbo Fisher (8-5)"
##                               2020                               2021
## "Jimbo Fisher (9-1)"      "Jimbo Fisher (8-4)"
```

Functions: do-it-yourself

```
# my function: input = number of wins; output ??  
jimbo.summary <- function(x){  
  total_w <- sum(x)  
  avg_w <- mean(x)  
  most_w <- max(x)  
  out <- c(total_w,avg_w,most_w)  
  names(out) <- c("total wins","avergae # wins","most wins")  
  return(out)  
}
```

My *jimbo* function: the output

```
# a vector with Jimbo's number of wins in Aggieland
jimbo <- c(8,9,8,9)

# Run the function
jimbo.summary(jimbo)
```

```
##      total wins avergae # wins      most wins
##          34.0           8.5           9.0
```

Our Aggie data

```
names(ags)

## [1] "Year"      "Wins"       "Losses"     "Pct"        "Coaching"   "Bowl"

nrow(ags)

## [1] 119

ncol(ags)

## [1] 6

dim(ags)

## [1] 119    6
```

Our Aggie data

```
summary(ags)
```

```
##      Year        Wins       Losses       Pct
## Min.  :1903   Min.   : 0.000   Min.   :0.000   Min.   :0.0500
## 1st Qu.:1932  1st Qu.: 4.000   1st Qu.:2.000   1st Qu.:0.4550
## Median :1962  Median : 6.000   Median :4.000   Median :0.6360
## Mean   :1962  Mean   : 6.252   Mean   :4.017   Mean   :0.6058
## 3rd Qu.:1992  3rd Qu.: 8.000   3rd Qu.:6.000   3rd Qu.:0.7640
## Max.   :2021  Max.   :12.000   Max.   :9.000   Max.   :1.0000
##      Coaching      Bowl
## Length:119      Length:119
## Class :character Class :character
## Mode  :character Mode  :character
##
##
```

Working with datafiles: Indexing

```
ags[1:5, "Wins"]

## # A tibble: 5 x 1
##   Wins
##   <dbl>
## 1     8
## 2     9
## 3     8
## 4     9
## 5     7
ags[c(1:5),]

## # A tibble: 5 x 6
##   Year Wins Losses Pct Coaching          Bowl
##   <dbl> <dbl>  <dbl> <dbl> <chr>
## 1 2021     8      4 0.667 Jimbo Fisher (8-4) Gator Bowl-Did ~
## 2 2020     9      1 0.9   Jimbo Fisher (9-1)  Orange Bowl-W
## 3 2019     8      5 0.615 Jimbo Fisher (8-5)  Texas Bowl-W
## 4 2018     9      4 0.692 Jimbo Fisher (9-4)  Gator Bowl-W
## 5 2017     7      6 0.538 Jeff Banks (0-1), Kevin Sumlin (7-5) Belk Bowl-L
```

Working with data: the \$ sign and Indexing

```
ags$Coaching[seq(from = 1, to = nrow(ags), by = 3)]  
  
## [1] "Jimbo Fisher (8-4)"      "Jimbo Fisher (9-4)"  
## [3] "Kevin Sumlin (8-5)"     "Kevin Sumlin (11-2)"  
## [5] "Mike Sherman (6-7)"     "Dennis Franchione (9-4)"  
## [7] "Dennis Franchione (4-8)" "R.C. Slocum (7-5)"  
## [9] "R.C. Slocum (9-4)"       "R.C. Slocum (10-0-1)"  
## [11] "R.C. Slocum (10-2)"     "Jackie Sherrill (7-5)"  
## [13] "Jackie Sherrill (10-2)"  "Jackie Sherrill (5-6)"  
## [15] "Tom Wilson (6-5)"       "Emory Bellard (10-2)"  
## [17] "Emory Bellard (5-6)"    "Gene Stallings (2-9)"  
## [19] "Gene Stallings (7-4)"   "Hank Foldberg (1-9)"  
## [21] "Jim Myers (4-5-1)"     "Jim Myers (4-6)"  
## [23] "Bear Bryant (7-2-1)"   "Ray George (3-6-1)"  
## [25] "Harry Stiteler (1-8-1)" "Homer Norton (4-6)"  
## [27] "Homer Norton (7-2-1)"  "Homer Norton (9-1)"  
## [29] "Homer Norton (5-2-2)"  "Homer Norton (2-7-2)"  
## [31] "Madison Bell (7-3)"    "Dana Bible (5-4-1)"  
## [33] "Dana Bible (7-1-1)"   "Dana Bible (5-4)"  
## [35] "Dana Bible (10-0)"    "E.H. Harlan (6-3)"  
## [37] "Charley Moran (3-4-2)" "Charley Moran (8-1)"  
## [39] "L.L. Larson (6-1-1)"   "J.E. Platt (4-2)"
```

Working with data: missing values

	Year	Wins	Losses	Pct	Coaching	Bowl
25	1996	6	6	0.500	R.C. Slocum (6-6)	NA
26	1995	9	3	0.750	R.C. Slocum (9-3)	Alamo Bowl-W
27	1994	10	0	0.955	R.C. Slocum (10-0-1)	NA
28	1993	10	2	0.833	R.C. Slocum (10-2)	Cotton Bowl-L
29	1992	12	1	0.923	R.C. Slocum (12-1)	Cotton Bowl-L
30	1991	10	2	0.833	R.C. Slocum (10-2)	Cotton Bowl-L
31	1990	9	3	0.731	R.C. Slocum (9-3-1)	Holiday Bowl-W
32	1989	8	4	0.667	R.C. Slocum (8-4)	Sun Bowl-L
33	1988	7	5	0.583	Jackie Sherrill (7-5)	NA
34	1987	10	2	0.833	Jackie Sherrill (10-2)	Cotton Bowl-W
35	1986	9	3	0.750	Jackie Sherrill (9-3)	Cotton Bowl-L
36	1985	10	2	0.833	Jackie Sherrill (10-2)	Cotton Bowl-W
37	1984	6	5	0.545	Jackie Sherrill (6-5)	NA
38	1983	5	5	0.500	Jackie Sherrill (5-5-1)	NA

How to deal with NAs?

```
# create vector of values 1-10, add NA to it
mis_vec <- c(1:10,NA)
mis_vec
```

```
## [1] 1 2 3 4 5 6 7 8 9 10 NA
```

```
# calculate mean of vector
mean(mis_vec)
```

```
## [1] NA
```

```
# better...
```

```
mean(mis_vec, na.rm = TRUE)
```

```
## [1] 5.5
```

Saving

Coding:

- ▶ We use script files - reproducing code.
- ▶ Save with the menu / disk sign on RStudio.
- ▶ File extension (name.R) is saved in your preferred directory.

Data files:

- ▶ If we changed the data, we can save the edited file.
- ▶ Use menu (save as...).
- ▶ Code: `write.csv(file_name, "selected_name.csv")`
- ▶ Data is saved in your preferred directory.

Packages



- ▶ Essential component of programming in R.
- ▶ User-generated 'stacks' of code/data.

- ▶ Free to download.
- ▶ Must be uploaded prior to use:
 - ▶ use the *library(package_name)* command.

Wrapping up week 1

Summary:

- ▶ What is Bush631?
- ▶ Why do I need to learn stats and research methods?
- ▶ Syllabus 'deep dive'.
- ▶ Intro to R: objects, vectors, functions, using data.

Homework assignments

Swirl tasks:

- ▶ Short practice of using R.
- ▶ Completed in RStudio **console**.
- ▶ Not sure how to answer? check the book.
- ▶ Submit **lessons 1-3 before** next class:
 1. Basic Building Blocks.
 2. Workspace and Files
 3. Sequences of Numbers

Swirl task submit - how?

For each lesson ↓