Bush 631-607: Quantitative Methods

Lecture 1 (08.31.2021): Introduction

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

Fall 2021

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.

- More? Check my website (Link)

Bush 631: What are we doing here?

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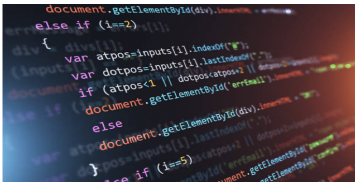
Quantitative Social Science

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

Studying global issues with data

Views of economy related to views of how country has handled COVID-19

% who say the current economic situation is __good__, among those who say (survey public) has done a ____ handling the coronavirus outbreak

Handling of COVID-19 outbreak:

	Bad job	Good job	Diff
Germany	41	79	+38
Canada	24%	62%	+38
Greece	6	42	+36
UK	23	55	+32
Belgium	33	62	+29
South Korea	7	36	+29
U.S.	19	42	+23
Sweden	70	92	+22
Netherlands	58	77	+19
France	18	35	+17
Spain	6	22	+16
Japan	13	27	+14
Italy	6	17	+11
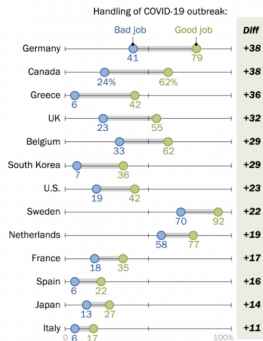
Note: All differences shown are statistically significant.
Source: Spring 2021 Global Attitudes Survey. Q1.
"Economic Attitudes Improve in Many Nations Even as Pandemic Endures"

PEW RESEARCH CENTER

Younger people more favorable toward UN, WHO's handling of COVID-19 and international cooperation
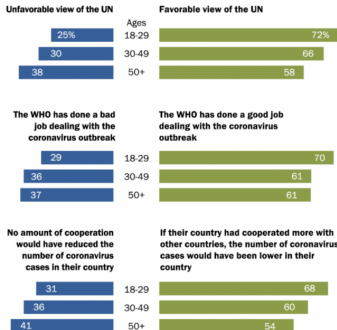
Median % who say ...

Unfavorable view of the UN	Ages	Favorable view of the UN
25%	18-29	72%
30	30-49	66
38	50+	58

The WHO has done a bad job dealing with the coronavirus outbreak		The WHO has done a good job dealing with the coronavirus outbreak
29	18-29	70
36	30-49	61
37	50+	61

| No amount of cooperation would have reduced the number of coronavirus cases in their country		If their country had cooperated more with other countries, the number of coronavirus cases would have been lower in their country
31	18-29	68
36	30-49	60
41	50+	54

Note: Percentages are medians based on each age group across 14 countries. "WHO" refers to the World Health Organization. In Australia and Canada, question asked about "COVID-19." In Japan, asked about "novel coronavirus." In South Korea, asked about "Corona19."
Source: Summer 2020 Global Attitudes Survey. Q4, Q8e, Q10e.
"International Cooperation Welcomed Across 14 Advanced Economies"

PEW RESEARCH CENTER

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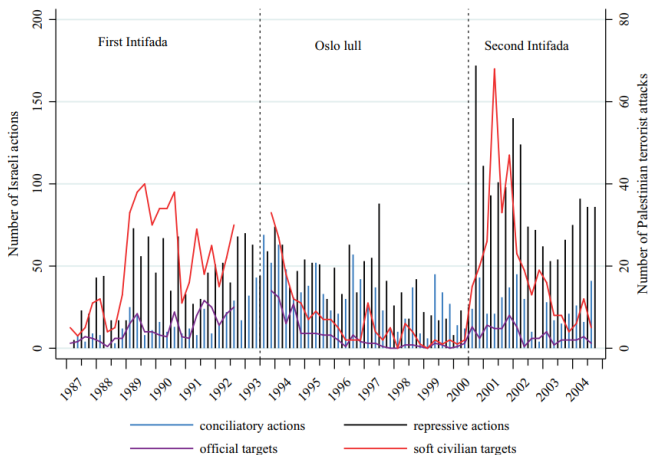
- ▶ What drives the likelihood of international conflicts?
- ▶ The leaders? or the regimes?
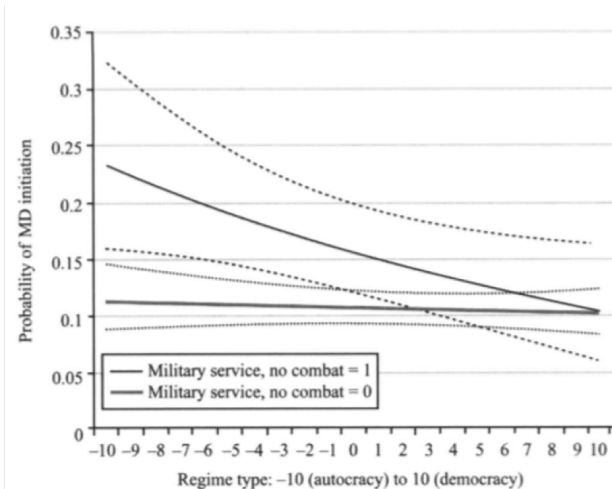
International Affairs research

▶ Counter terrorism strategies and organizational responses.

| conciliatory actions | repressive actions |
| official targets | soft civilian targets |

International Affairs research

- Diplomacy, political economy (trade, sanctions, foreign aid).
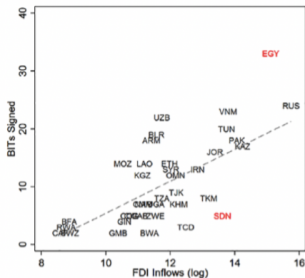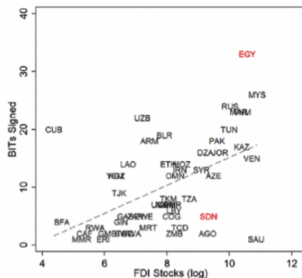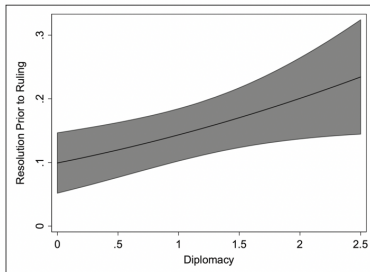
Why should I take this class?

- 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.

- You have to? :)

How are we doing it?

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

Syllabus 'deep dive'

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

Syllabus 'deep dive'

The essentials:

- When: Tuesdays, 4:35-7:20pm
- Where: Allen 1006

Office hours:

- Tuesday & Thursday 9:30-11:00am.
- My office: Allen 2116.
- 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: class tasks.

Course material and resources:

- ▶ Website: lecture slides, R code, other resources (Soon).
- ▶ 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 if you miss class.

COVID is still here...

- Please don't come to class if you're sick/tested positive.
- Follow A&M guidelines (forms, quarantine).
- University accepted excused absence.
- If you've been a close contact - share with us please!

Syllabus 'deep dive'

Grading and assignments:

1. Attendance (5%).
2. Home assignments (Swirl - R) (10%).
3. Research design tasks - 3 tasks (40%).
4. Final project - policy brief (35%).
5. Final project presentation (10%).
6. Xtra credit - TBD

Syllabus 'deep dive'

Other issues:

- ▶ Make-up policy.
- ▶ Plagiarism.
- ▶ COVID: prioritize your and others' health.
- ▶ Electronics in class.
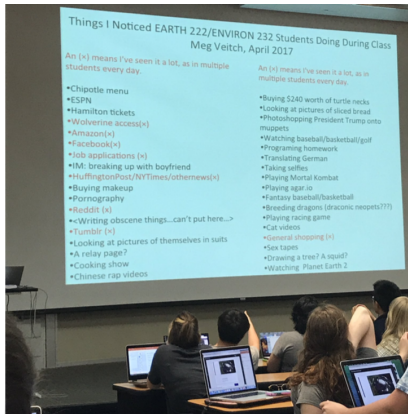
Syllabus 'deep dive'

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
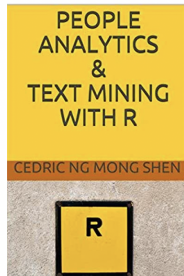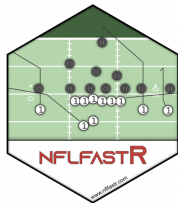
Introduction to R

Syllabus Task with R

Bush 631: Sports teams preferences

Syllabus Task with R

2010s Dominance?

Titles since 2010 / Sports

Bayern Munich, Ronaldo, USWNT, Durant, Bregman, Spurs, Chiefs, Rousey, Westbrook, Watt, Stars, Baines

MLB, NBA, NFL, NHL, Soccer, UFC

Bush 631: Sports teams origins

US-based sport, Non-US sports

Proportions
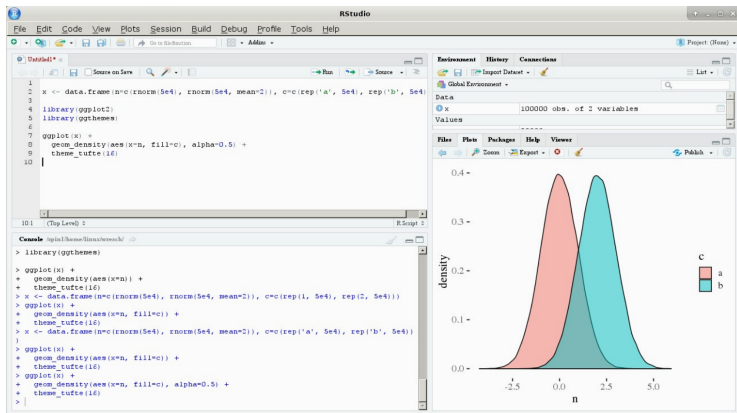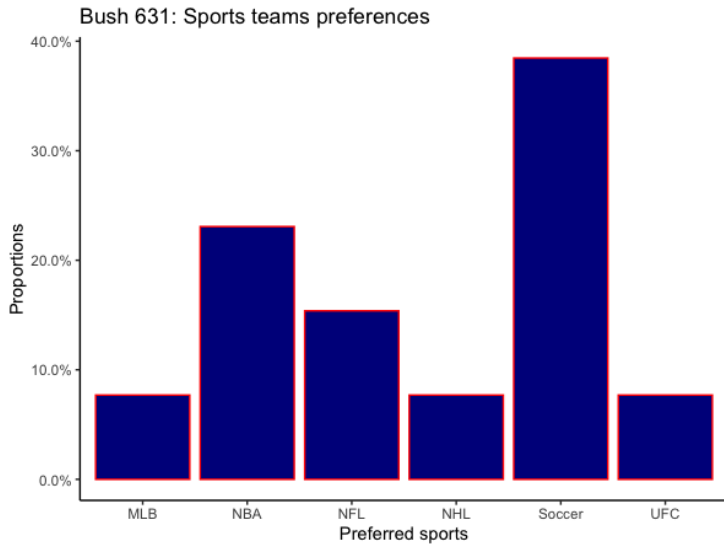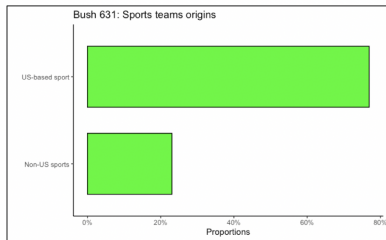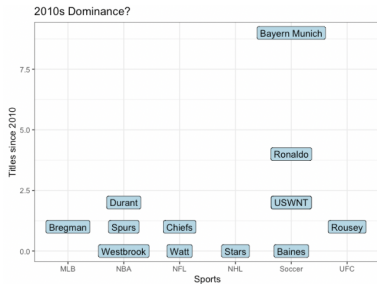
Syllabus Task with R

Bush 631: Sports preferences (some more info)

Team/Athlete names	Sports	Gender Diversity	Texas Based
Bregman	MLB	Men	Yes
Durant	NBA	Men	No
Spurs	NBA	Men	Yes
Westbrook	NBA	Men	No
Chiefs	NFL	Men	No
Watt	NFL	Men	Yes
Stars	NHL	Men	Yes
Baines	Soccer	Men	No
Bayern Munich	Soccer	Men	No
Ronaldo	Soccer	Men	No
USWNT	Soccer	Women	No
USWNT	Soccer	Women	No
Rousey	UFC	Women	No

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

| Environment | History | Connections |
| Import Dataset ▾ | List ▾ |

Data
| From Text (base)... |
| From Text (readr)... |
| From Excel... |
| From SPSS... |
| From SAS... |
| From Stata... |

annu... ables
congr... ariables
dem variables
dem11 iables
dem80 iables
gulf.out 217 obs. of 70 variables
GulfData 217 obs. of 358 variables
mydata 9120 obs. of 6 variables
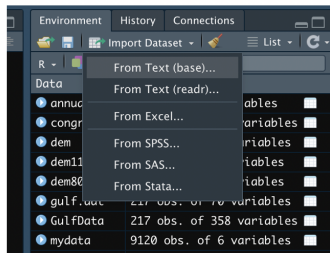
2) Direct code

```
126  setwd("~/Dropbox/TAMU/Bush631_QuantMethods")
127
128  library(readxl)
129  ags <- read_excel("ags.xlsx")
130
131  View(ags)
```

(a) Define working directory
(b) Use the read_ function
(c) View – opens data file

Our data

	Year	Wins	Losses	Pct	Coaching	Bowl
1	2020	9	1	0.900	Jimbo Fisher (9–1)	Orange Bowl–W
2	2019	8	5	0.615	Jimbo Fisher (8–5)	Texas Bowl–W
3	2018	9	4	0.692	Jimbo Fisher (9–4)	Gator Bowl–W
4	2017	7	6	0.538	Jeff Banks (0–1), Kevin Sumlin (7–5)	Belk Bowl–L
5	2016	8	5	0.615	Kevin Sumlin (8–5)	Texas Bowl–L
6	2015	8	5	0.615	Kevin Sumlin (8–5)	Music City Bowl–L
7	2014	8	5	0.615	Kevin Sumlin (8–5)	Liberty Bowl–W
8	2013	9	4	0.692	Kevin Sumlin (9–4)	Chick-fil-A Bowl–W
9	2012	11	2	0.846	Kevin Sumlin (11–2)	Cotton Bowl–W
10	2011	7	6	0.538	Mike Sherman (6–6), Tim DeRuyter (1–0)	Meineke Car Care Bowl of Texas–W
11	2010	9	4	0.692	Mike Sherman (9–4)	Cotton Bowl–L
12	2009	6	7	0.462	Mike Sherman (6–7)	Independence Bowl–L
13	2008	4	8	0.333	Mike Sherman (4–8)	*NA*
14	2007	7	6	0.538	Dennis Franchione (7–6)	Alamo Bowl–L
15	2006	9	4	0.692	Dennis Franchione (9–4)	Holiday Bowl–L
16	2005	5	6	0.455	Dennis Franchione (5–6)	*NA*

Working with data: **Indexing**

```
ags[1,]
```

```
## # A tibble: 1 x 6
##    Year  Wins Losses   Pct Coaching          Bowl
##   <dbl> <dbl>  <dbl> <dbl> <chr>             <chr>
## 1  2020     9      1   0.9 Jimbo Fisher (9-1) Orange Bowl-W
```

```
ags[,1]
```

```
## # A tibble: 118 x 1
##     Year
##    <dbl>
##  1  2020
##  2  2019
##  3  2018
##  4  2017
##  5  2016
##  6  2015
##  7  2014
##  8  2013
##  9  2012
## 10  2011
## # ... with 108 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  2020     9      1 0.9   Jimbo Fisher (9-1)               Orange Bowl-W
## 2  2019     8      5 0.615 Jimbo Fisher (8-5)               Texas Bowl-W
## 3  2017     7      6 0.538 Jeff Banks (0-1), Kevin Sumlin (7-5) Belk Bowl-L
```

```
ags[1:3]
```

```
## # A tibble: 118 x 3
##     Year  Wins Losses
##    <dbl> <dbl>  <dbl>
## 1   2020     9      1
## 2   2019     8      5
## 3   2018     9      4
## 4   2017     7      6
## 5   2016     8      5
## 6   2015     8      5
## 7   2014     8      5
## 8   2013     9      4
## 9   2012    11      2
## 10  2011     7      6
## # ... with 108 more rows
```

Working with data: **Indexing**

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

```
## # A tibble: 118 x 1
##    Coaching
##    <chr>
##  1 Jimbo Fisher (9-1)
##  2 Jimbo Fisher (8-5)
##  3 Jimbo Fisher (9-4)
##  4 Jeff Banks (0-1), Kevin Sumlin (7-5)
##  5 Kevin Sumlin (8-5)
##  6 Kevin Sumlin (8-5)
##  7 Kevin Sumlin (8-5)
##  8 Kevin Sumlin (9-4)
##  9 Kevin Sumlin (11-2)
## 10 Mike Sherman (6-6), Tim DeRuyter (1-0)
## # ... with 108 more rows
```

```
ags[1:10, c("Year","Bowl")]
```

```
## # A tibble: 10 x 2
##    Year Bowl
##    <dbl> <chr>
##  1  2020 Orange Bowl-W
##  2  2019 Texas Bowl-W
##  3  2018 Gator Bowl-W
##  4  2017 Belk Bowl-L
##  5  2016 Texas Bowl-L
##  6  2015 Music City Bowl-L
##  7  2014 Liberty Bowl-W
##  8  2013 Chick-fil-A Bowl-W
##  9  2012 Cotton Bowl-W
## 10  2011 Meineke Car Care Bowl of Texas-W
```

Working with data: *using the $ sign*

```
ags$Year[5]
```

```
## [1] 2016
```

```
ags$Coaching[1:5]
```

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

Math operations and data vectors

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

ags_win_p
```

```
##   [1]  90.0  61.5  69.2  53.8  61.5  61.5  61.5  69.2  84.6  53.8  69.2  46.2
##  [13]  33.3  53.8  69.2  45.5  58.3  33.3  50.0  66.7  58.3  66.7  78.6  69.2
##  [25]  50.0  75.0  95.5  83.3  92.3  83.3  73.1  66.7  58.3  83.3  75.0  83.3
##  [37]  54.5  50.0  45.5  58.3  36.4  54.5  66.7  66.7  83.3  83.3  72.7  45.5
##  [49]  27.3  45.5  18.2  30.0  30.0  63.6  45.0  30.0  10.0  25.0  30.0  45.0
##  [61]  25.0  30.0  40.0  72.7  95.0  75.0  10.0  45.0  35.0  60.0  63.6  15.0
##  [73]   5.0  35.0  40.0  60.0  63.6  75.0  45.0  81.8  90.0 100.0  50.0  66.7
##  [85]  70.8  30.0  27.3  65.0  50.0  70.0  22.2  55.6  55.0  94.4  61.1  83.3
##  [97]  75.0  61.1  55.6  77.8  81.3 100.0  85.7 100.0  66.7  75.0  81.3  44.4
## [109]  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] 6.237288
mean(ags$Pct)
```

```
## [1] 0.6052966
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] 6.237288
```

Functions: code examples

```r
sec <- seq(from = 2012, to = 2020, by = 1)
sec_coach <- ags$Coaching[9:1]
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
##           "Jimbo Fisher (9-1)"
```

Functions: do-it-yourself

```r
# 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(9,8,9)

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

```
##       total wins avergae # wins      most wins
##        26.000000       8.666667       9.000000
```

Our Aggie data

```
names(ags)
```

```
## [1] "Year"     "Wins"     "Losses"   "Pct"      "Coaching" "Bowl"
nrow(ags)
```

```
## [1] 118
ncol(ags)
```

```
## [1] 6
dim(ags)
```

```
## [1] 118   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 :0.6255
## Mean   :1962   Mean   : 6.237   Mean   :4.017   Mean   :0.6053
## 3rd Qu.:1991   3rd Qu.: 8.000   3rd Qu.:6.000   3rd Qu.:0.7710
## Max.   :2020   Max.   :12.000   Max.   :9.000   Max.   :1.0000
##    Coaching           Bowl
## Length:118        Length:118
## Class :character   Class :character
## Mode  :character   Mode  :character
##
##
##
```

Working with datafiles: **Indexing**

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

```
## # A tibble: 5 x 1
##     Wins
##    <dbl>
## 1      9
## 2      8
## 3      9
## 4      7
## 5      8
```

```
ags[c(1:5),]
```

```
## # A tibble: 5 x 6
##    Year  Wins Losses   Pct Coaching                          Bowl
##   <dbl> <dbl>  <dbl> <dbl> <chr>                             <chr>
## 1  2020     9      1 0.9   Jimbo Fisher (9-1)                Orange Bowl-W
## 2  2019     8      5 0.615 Jimbo Fisher (8-5)                Texas Bowl-W
## 3  2018     9      4 0.692 Jimbo Fisher (9-4)                Gator Bowl-W
## 4  2017     7      6 0.538 Jeff Banks (0-1), Kevin Sumlin (7-5) Belk Bowl-L
## 5  2016     8      5 0.615 Kevin Sumlin (8-5)                Texas Bowl-L
```

Working with data: the $ sign and Indexing

```
ags$Coaching[seq(from = 1, to = nrow(ags), by = 3)]
```

```
##  [1] "Jimbo Fisher (9-1)"
##  [2] "Jeff Banks (0-1), Kevin Sumlin (7-5)"
##  [3] "Kevin Sumlin (8-5)"
##  [4] "Mike Sherman (6-6), Tim DeRuyter (1-0)"
##  [5] "Mike Sherman (4-8)"
##  [6] "Dennis Franchione (5-6)"
##  [7] "R.C. Slocum (6-6)"
##  [8] "R.C. Slocum (8-4)"
##  [9] "R.C. Slocum (6-6)"
## [10] "R.C. Slocum (10-2)"
## [11] "R.C. Slocum (9-3-1)"
## [12] "Jackie Sherrill (10-2)"
## [13] "Jackie Sherrill (6-5)"
## [14] "Tom Wilson (7-5)"
## [15] "Emory Bellard (4-2), Tom Wilson (4-2)"
## [16] "Emory Bellard (10-2)"
## [17] "Emory Bellard (3-8)"
## [18] "Gene Stallings (3-7)"
## [19] "Gene Stallings (4-5-1)"
## [20] "Hank Foldberg (2-7-1)"
## [21] "Jim Myers (1-6-3)"
## [22] "Bear Bryant (8-3)"
## [23] "Bear Bryant (1-9)"
```

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?

```r
# 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
```

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

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

```r
# 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 working directory.

Data files:

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

Packages

- ▶ Essential component of programming in R.
- ▶ User-generated 'stacks' of code/data.
- ▶ Free to download.
- ▶ Must be uploaded prior to use: using the library(package_name) command.

Wrapping up week 1

Busy today:

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

Lecture 1 full code: website/Canvas

Home assignments

Swirl tasks:

- ▶ Short practice of using R.
- ▶ Completed in RStudio.
- ▶ Not sure how to answer? check the book.
- ▶ Submit task: INTRO1 **before** next class.

Swirl task submit - how?