

POLS 2000 Webbook 2.0

Bill Wilkerson

2023-07-26

Table of contents

Preface	3
I GETTING READY TO WORK	4
1 RStudio Cloud	5
1.1 Accessing RStudio Cloud	5
1.2 Accessing and Using Assignments	5
1.3 Importing Files into an Assignment	5
1.4 Exporting Files from an Assignment	5
2 Navigating RStudio	6
2.1 The RStudio Layout	6
2.2 Customize How RStudio Looks	7
3 Using Quarto Documents	9
3.1 What is a Quarto Document?	9
3.1.1 Creating a Quarto Document	9
3.2 Working in Quarto Documents	10
3.2.1 The Visual Editor vs. the Source Editor	10
3.2.2 Writing in Markdown	11
3.2.3 Code Chunks	11
3.3 Rendering Quarto Documents	13
3.4 Exporting Quarto Documents	13
3.5 Learning More About Quarto	13
II YOUR DATASET	15
4 Choosing Your Dataset	16
4.1 What is a Dataset?	16
4.2 Three Datasets to Choose From	16
4.3 Connecting Your Research Question to Data: Codebooks	16

III SKILLS & TECHNIQUES	17
5 Loading and Saving Data	18
5.1 Loading Data	18
5.2 Saving Data	18
5.2.1 A Note on Naming Datasets and Files	19
6 Managing Data	20
6.1 Select	21
6.2 Filter	22
6.3 Mutate	22
7 Graphs for One Variable	23
8 Statistics for One Variable	24
9 Graphs for Two or More Variables	25
10 Tables for 2 or More Variables	26
11 Hypothesis Testing	27
12 Regression Analysis	28
12.1 Logistic Regression	28
12.2 Linear Regression	28
Appendices	29
SA 2: 2022 ANES Pilot Study	29
Defining your research question, dataset, & variables	29
Research Question	29
Dataset Selected	29
Variables Selected for Initial Study	29
Load packages	30
Load the correct dataset	30
Using the dataset you have loaded, select your variables, and save your new smaller dataset	30
Run variable frequencies on your new smaller dataset	30
SA 2: WVS	33
Defining your research question, dataset, & variables	33
Research Question	33
Dataset Selected	33

Variables Selected for Initial Study	33
Load packages and run function for selecting a country	34
Select a country to study, select variables, and save your new smaller dataset (link to help)	34
Run variable frequencies on your new smaller dataset	35
SA 2: ANES 2020	39
Defining your research question, dataset, & variables	39
Research Question	39
Dataset Selected	39
Variables Selected for Initial Study	39
Load packages	40
Load the correct dataset	40
Using the dataset you have loaded, select your variables, and save your new smaller dataset	40
Run variable frequencies on your new smaller dataset	41
SA 3: Data Management Strategies	45
Discussion	45
Load Packages	45
Load Your Datasets	45
Select variables	46
Filter Your Dataset	46
Recode a Variable	46
Create Another New Variable	46
Run Frequencies for the New Variables You Created	47
Save your updated dataset?	49
SA 4: Single Variable EDA	50
Discussion	50
Load Packages	50
Load Your Dataset	50
Manage your data as needed	50
Graph 1	51
Graph 2	51
Graph 3 (add more if you want to!)	52
Run Univariate Statistics	53
Save your updated dataset?	59
SA 6: Two or more variable EDA	60
Discussion	60
Load Packages	60
Load Your Dataset	61

Manage your data as needed	61
Graph & Cross-tab 1	61
Graph & Cross-tab 2	62
Graph & Cross-tab 3 (add more if you want to!)	64
.	65
Save your updated dataset?	65
SA 7: Hypothesis Testing	66
Discussion	66
Load Packages	66
Load Your Dataset	66
Manage your data as needed	66
Hypothesis Test 1	67
Hypothesis Test 2	68
Hypothesis Test 3 (add more if needed!)	70
.	71
Save your updated dataset?	71
SA 8: Logistic Regression Analysis	72
Discussion	72
Load Packages	72
Load Your Dataset	72
Manage your data as needed	72
Final Regression Model	73
Save your updated dataset?	75
SA 8: Linear Regression Analysis	76
Discussion	76
Load Packages	76
Load Your Dataset	76
Manage your data as needed	76
Final Regression Model	77
Save your updated dataset?	79

Preface



This semester you are conducting a quantitative project using an NSF-funded curriculum, [Passion Driven Statistics](#). This is a set of tutorials—text and video—to support your work. My goal has been to keep these tutorials as simple as possible. You may need to go beyond these tutorials to conduct your research. I am here to help. Along the way you will learn to:

- manage your data
- explore your data through graphs (or in R terms plot your data)
- test hypotheses using your data

This is a first draft. I ask that as you use these tutorials, you will note errors and suggest improvements.

In addition to Passion Driven Statistics, these tutorials borrow from the following:

- [Data Carpentry](#)
- [Modern Dive](#)
- [The PowerMut Project](#)
- [RLadies Sydney](#)

And a big thanks to Ed Beck of the SUNY Oneonta TLTC for getting this project moving and for all of his good advice—that I sometimes ignore to my detriment—since the pandemic began.



Part I

GETTING READY TO WORK

1 RStudio Cloud

complete this file once a decision is made

1.1 Accessing RStudio Cloud

To be added.

1.2 Accessing and Using Assignments

To be added.

1.3 Importing Files into an Assignment

To be added.

1.4 Exporting Files from an Assignment

To be added.

2 Navigating RStudio

2.1 The RStudio Layout

The first time you open an assignment in RStudio Cloud, the console window will take up the entire left side. You will see three panes.

Once you open a script file, you will see the four-quadrant layout seen below.

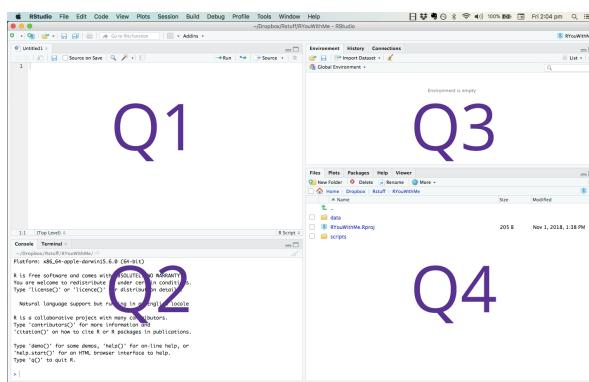


Figure 2.1: RStudio Four Tile Layout

Described below are the features in the four quadrants.

- **Q1. The Script Editor window** (upper-left quadrant above) is a workspace where you can write, edit, and save R commands. Rather than entering these commands in the console window, you can run them from the script editor. The advantage is that you can easily edit, save and re-run all your analyses.
- **Q2. The Console window** (lower-left quadrant above) displays results. You can interact with R directly in the console window by typing commands. Use the script editor, though, since you can save your work for later use.
- **Q3. The Environment Window** (upper-right quadrant above) gives you a list of all the objects you have created, including datasets, graphs, and statistical model results.
- **Q4. The Files window** (lower-right quadrant above) provides a *lot* of information. The **files** tab gives you a file explorer. The **plots** tab provides a history of the graphs you have made. The **packages** tab shows you the packages you have available and the ones you have enabled. The **help** tab allows you to see help files for functions and packages. Occasionally, things you are trying to see will show up in the **viewer** tab.

Note: the size of each of these quadrants is adjustable by dragging the edges.

2.2 Customize How RStudio Looks

You can customize how R Studio looks to make it work for you. Click Tools - Global Options - Appearance. You can change the default font, font size, and general appearance of R Studio, including using dark mode. Play around with the settings and see what you prefer. You're going to spend a lot of time with R, it might as well look nice!

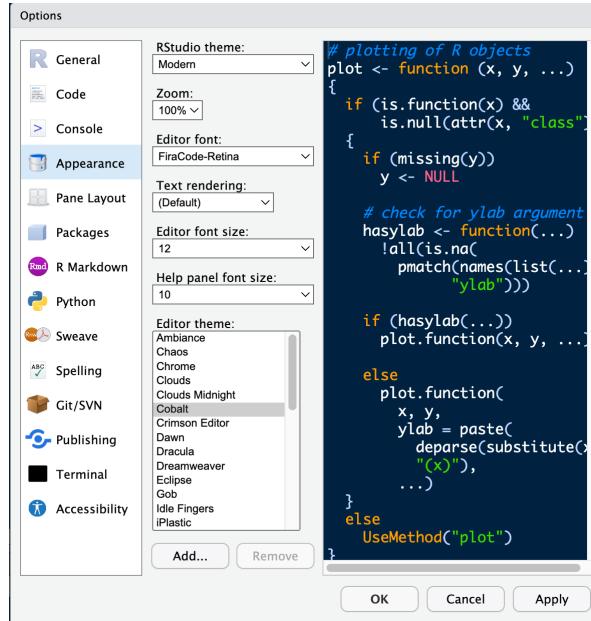


Figure 2.2: RStudio Four Tile Layout

This is the layout I use on my computer.

There are a lot of options! But don't let picking a theme get in the way of getting work done.



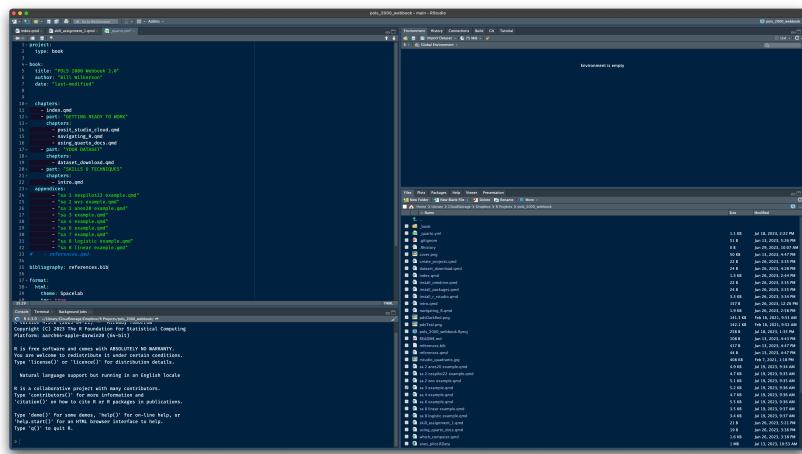


Figure 2.3: RStudio Four Tile Layout

3 Using Quarto Documents

3.1 What is a Quarto Document?¹

This semester you will be using the statistical programming language R in RStudio, an integrated development environment (or IDE). But where will we put the code? We will use an interactive document, Quarto. Quarto documents can contain both code, that we can execute to run data analysis and create graphs, and text that discusses the work we have done.

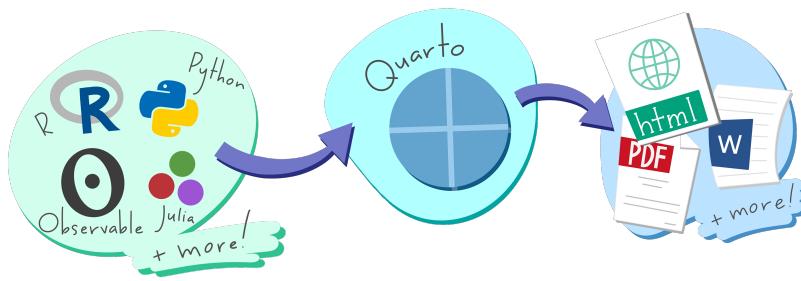


Figure 3.1: Quarto incorporates R code and text and can produce documents in many formats.

Quarto documents are interactive in that we can run our analysis and then write up explanations of it. Quarto documents can be rendered, running the code and producing a report in a specific format such as Word, .PDF, .HTML, and many other formats.

3.1.1 Creating a Quarto Document

We will mostly be working with templates that you can work to fill in. But creating a Quarto document is easy. Starting and Naming a New Quarto Document Start a new Quarto document in RStudio by clicking File > New File > Quarto Document...

¹Much of this section has been used under Creative Commons license from the Data Carpentry course, [Introduction to Reproducible Publications in Rstudio](#).

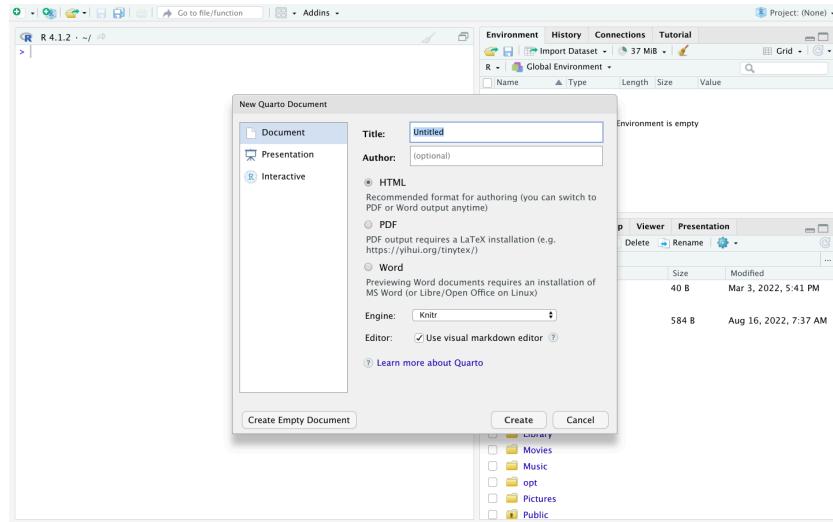


Figure 3.2: Opening a new Quarto document.

You may name your Quarto document as “my_first_first_qmd”.

New Quarto files will have a generic template unless you click the “Create Empty Document” in the bottom left-hand corner of the dialog box.

We will keep all pre-selected options: HTML as the output, knitr engine and the visual editor. The output might be changed at anytime and we can easily switch between the visual and the source editor. Knitr will be the engine used to execute the R codes and to render the document in Rstudio.

If you see this default text you’re good to go:

3.2 Working in Quarto Documents

You can work in two modes: the visual editor or the source editor. Most of you will want to work in visual editor mode.

3.2.1 The Visual Editor vs. the Source Editor

The visual editor follows the WYSIWYG “what you see is what you get” approach similarly to Word or Google docs that lets you choose styling options from the menu (before you had to either have the markdown code memorized or look it up for each of your styling choices). Another major benefit is that the new editor renders the styling in real time so you can preview

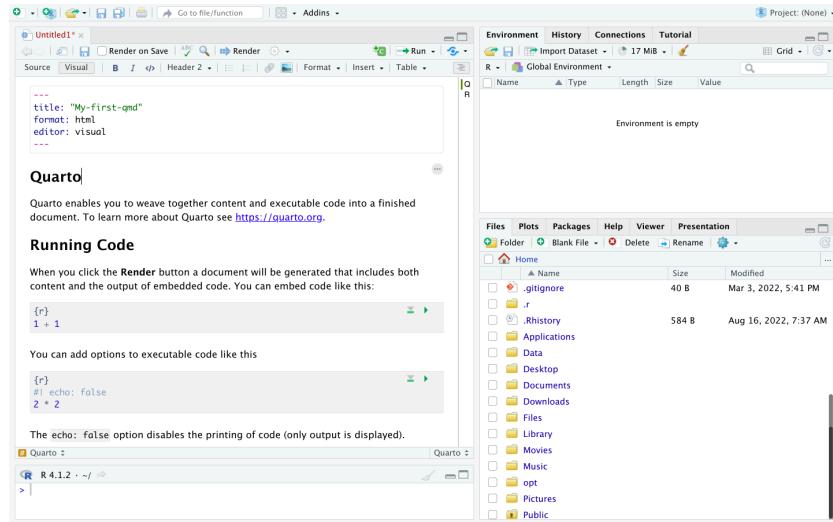


Figure 3.3: A new Quarto document in visual editor mode.

your paper before rendering to your output format. The image directly above shows the visual editor.

In the visual editor you can use editor tool bar at the top of the document to format your document. Here are the options.

If you toggle the source button, you will display your Quarto document in the “source editor” mode. Notice the symbols scattered throughout the text (#, *, <>). Those are examples of Markdown syntax, an easy and quick, human-readable markup language for document styling.

3.2.2 Writing in Markdown

If you prefer to write in the source editor, you will need to use a version of Markdown, called RMarkdown. Moving between the source and visual editor should help you get a sense of the code.

Here is [a short reference for Markdown code](#).

3.2.3 Code Chunks

Whether you are working in the visual or the source editor, you need to put place your R code in sections that are marked to run the code named “code chunks”.

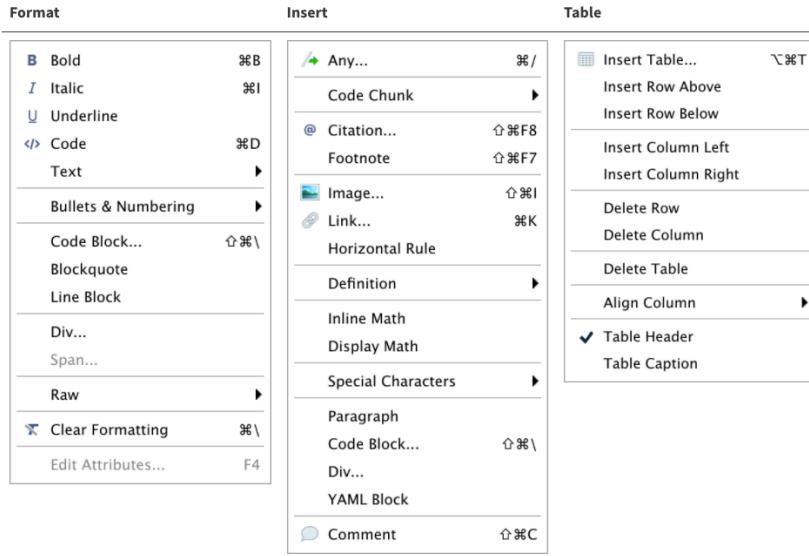


Figure 3.4: The Quarto visual editor toolbars.

The screenshot shows a Quarto document in the source editor. The code in the left pane is:

```

1 ---  
2 title: "My-First-qmd"  
3 format: html  
4 editor: visual  
5 ---  
6  
7 # Quarto  
8  
9 Quarto enables you to weave together content and executable code into a  
finished document. To learn more about Quarto see https://quarto.org.  
10  
11 # Running Code  
12  
13 When you click the **Render** button a document will be generated that  
includes both content and the output of embedded code. You can embed code  
like this:  
14  
15 ``{r}  
16 1 + 1  
17 ``  
18  
19 You can add options to executable code like this  
20  
21 ``{r}  
22 # echo: false  
23 2 * 2  
24 ``  
25  
26 The `echo: false` option disables the printing of code (only output is  
displayed).
    
```

The right pane shows the environment variables:

Name	Type	Length	Size	Value
Environment is empty				

The bottom pane shows the file system:

Files	Plots	Packages	Help	Viewer	Presentation
Quarto					
Home					
..					
Ignore					
.r					
Rhistory					
Applications					
Data					
Desktop					
Documents					
Downloads					
Files					
Library					
Movies					
Music					
opt					
Pictures					
Public					

Figure 3.5: A Quarto document in the source editor.

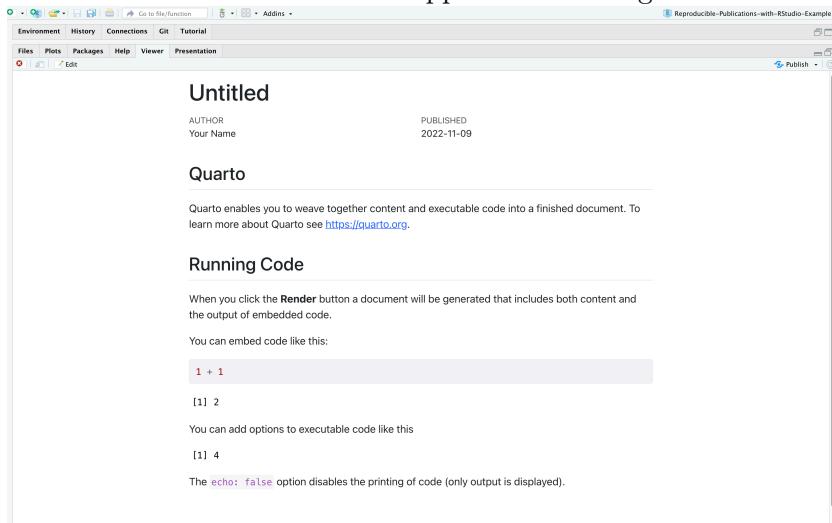
You will need to put everything you need to complete an assignment in a code chunk. This includes loading packages and data, and the code for the analyses and graphs you want to run.

You can run individual code chunks by clicking on the green triangles in each code chunk.

3.3 Rendering Quarto Documents

Simply put, rendering is the process of converting a document into a file format or a medium that is paginated or has the concept of pages. Clicking the render button will compile the code, check for errors, and finally, output the type of file indicated in your yaml header. You may select the option "Render on Save" to see a preview of your document every time you save edits. Attention: your qmd document may not run and render as your indicated output if there are any errors in the document so it also functions somewhat as a code checker.

The rendered document will appear in the right side of RStudio in the Viewer area.



3.4 Exporting Quarto Documents

Complete once I know whether I am using on individual computers or RStudio Cloud.

3.5 Learning More About Quarto

- [Introduction to Reproducible Publications with RStudio](#)
- [Quarto Reference by Posit](#)

- R for Data Science

Part II

YOUR DATASET

4 Choosing Your Dataset

4.1 What is a Dataset?

4.2 Three Datasets to Choose From

You have a choice between three datasets to work with, this semester:

- The 2020 American National Election Study —
- The 2022 American National Election Study Pilot —
- The World Values Survey —

4.3 Connecting Your Research Question to Data: Codebooks

Part III

SKILLS & TECHNIQUES

5 Loading and Saving Data

5.1 Loading Data

You first need to place your dataset into the RStudio environment to do your work. The datasets you will work with are in a native-R format, `.RData`. To do this, you use the `load()` command. Make sure to put quotation marks around the file name inside the parentheses. Here is an example.

```
load("smaller_anes_2020.RData")
```

5.2 Saving Data

You can reload your dataset every time you use it, but it is simpler to save the dataset and load that file. Use the `save` command. Make sure that you use the name that is in your work space in the `save` command and you should make sure that both names are the same. You can rename your working file if needed. I have changed the name and then saved the file below.

```
name_in_workspace <- some_other_name  
save(name_in_workspace, file = "name_in_workspace.RData")
```

Using the Ethiopia survey in the World Values Survey dataset as an example your name could be:

```
save(ethiopia, file = "ethiopia.RData")
```

It is critical that you:

- put quotation marks around your file name,
- and use `.RData` as a file extension.

5.2.1 A Note on Naming Datasets and Files

As you make changes in your dataset during the semester, you will want to save your updated file. Give each version a new name so that you can go back to your previous work. How to name your files? Here are a few good practices.

- You cannot space in a file name.
- You cannot start a file name with a number.
- Make them easy to read. Use snake_case, underscores “_” between words.
- Make them easy to type, use all lower case.
- Don’t use words like “final”, rather number your files consistently.

So, if you are using the Ethiopia survey and make changes something like `ethiopia_2.RData` works.

6 Managing Data

The Passion Driven Statistics video calls this section of the course *data management*. I prefer *data wrangling*. You are trying to get control of your data so that you and your data can get work done.



Much of the work of anyone who works with quantitative data is data wrangling: creating a dataset for the particular task you have in mind. Work might include subsetting data, filtering data, creating new variables, or making the data easier to understand.

Much of the essential work has been done for you. The datasets you will start with are in R format, but you will need to do some data wrangling.

To do most of this work, you will use an R package, `{dplyr}`, part of the `{tidyverse}`, a group of packages created to make working with data more manageable. Specifically, the functions you will be using are:

- **select** — choose variables from a dataset to include or exclude
- **filter** — pick observations, rows, in a dataset by their values
- **mutate** — create new variables
- **arrange** — rearrange the rows in a dataset
- **summarize** — collapse many values into a single summary

These functions can be combined with other packages to wrangle data in impressive ways. This chapter will give examples of using these verbs for everyday data management tasks.

6.1 Select

Each of the datasets you can choose from are big, they you have datasets with *a lot* of variables. The World Values Survey has 487 variables. The American National Election Study datasets are even bigger. The computer can handle datasets this big, but they are unwieldy. You should choose a subset of variables you want to work with in your research this semester.

To subset a dataset, we'll use the `select` function in the `{dplyr}` package. `{dplyr}` is part of the `{tidyverse}`.

Here is a short video (about 3:00) on using the `select` done by an R user group, [R-Ladies Sydney](#):

Let's look at an example from the WVS. I have been using Ethiopia. For now, my research question is: Does where a person lives have an impact on Ethiopian's views on democracy? Looking at the codebook, I see a series of variables on where folks reside and a series of questions on democracy. So I am going to select only the variables I want to look at for now. We can always add to or subtract from this list later.

```
library(tidyverse)

ethiopia_small <- ethiopia |>
  select(N_REGION_WVS, N_TOWN, G_TOWNSIZE, G_TOWNSIZE2, H_SETTLEMENT, H_URBRURAL, Q34, Q24)
```

A lot of typing! Since these variables are in two groups that are consecutive in the dataset, I can do this by using a shortcut `:``. This shortcut selects all the variables in a range of consecutive variables. [There are a bunch of shortcuts, helpers, for the select function.](#)

```
library(tidyverse)

ethiopia_small <- ethiopia |>
  select(N_REGION_WVS : H_URBRURAL, Q34, Q241 : Q261, E1_LITERACY, Q275R)
```

One helper that I will show you is unselecting variables. Let's say you don't want include a couple of variables anymore. We can unselect them by putting the minus (-) sign before them in the select command.

```
ethiopia_small_2 <- ethiopia_small %>%
  select(-E1_LITERACY, -Q275R)
```

This drops two variables, `E1_LITERACY` and `Q275R` from our selected variables and creates a new dataset with a new name, `ethiopia_small_2`.

For more details on using filter see either XXX or YYY.



6.2 Filter

6.3 Mutate

7 Graphs for One Variable

8 Statistics for One Variable

9 Graphs for Two or More Variables

10 Tables for 2 or More Variables

11 Hypothesis Testing

12 Regression Analysis

12.1 Logistic Regression

12.2 Linear Regression

SA 2: 2022 ANES Pilot Study

Defining your research question, dataset, & variables

Find Quarto themes here.

Research Question

To what extent do socio-economic and other factors explain support or opposition to the 2022 *Dobbs* decision?

Dataset Selected

2022 ANES Pilot Study

Variables Selected for Initial Study

Table 1: Table 1. Variables Chosen for Study

Purpose of Variable	Variable Name & Description	Level of Measurement
Dependent	roe- favor or oppose overturning <i>Roe v. Wade</i>	nominal
Independent	urbanicity2 - urban-rural status pid3 — party identification of respondent ideo5 — ideology of respondent gender — Gender of respondent educ — education level of respondent pew_religimp — Importance of religion in your life follow - Do you follow what's going on in government and public affairs?	ordinal ordinal ordinal nominal ordinal ordinal ordinal

Purpose of Variable	Variable Name & Description	Level of Measurement

Load packages

```
library(tidyverse)
library(haven)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(ggblanket)
```

Load the correct dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("anes_pilot.RData")
```

Using the dataset you have loaded, select your variables, and save your new smaller dataset

Need help? [Go to chapter 4 in the webbook.](#)

```
anes_pilot_smaller <- anes_pilot |>
  select(roe, urbanicity2, pid3, ideo5, gender, pew_religimp, follow, caseid, weight)
```

Run variable frequencies on your new smaller dataset

Need help? [Go to chapter x in the webbook.](#)

```

anes_pilot_smaller |>
  select(-caseid, -weight) |>
  frq()

Favor/oppose - overturn Roe v. Wade (roe) <numeric>
# total N=1585 valid N=1585 mean=1.85 sd=0.71

Value | Label | N | Raw % | Valid % | Cum. %
-----
1 | Favor | 526 | 33.19 | 33.19 | 33.19
2 | Oppose | 764 | 48.20 | 48.20 | 81.39
3 | Neither favor nor oppose | 295 | 18.61 | 18.61 | 100.00
<NA> | <NA> | 0 | 0.00 | <NA> | <NA>

Profile: Urban-rural status (urbanicity2) <numeric>
# total N=1585 valid N=1585 mean=2.92 sd=1.33

Value | Label | N | Raw % | Valid % | Cum. %
-----
1 | Big city | 317 | 20.00 | 20.00 | 20.00
2 | Smaller city | 244 | 15.39 | 15.39 | 35.39
3 | Suburban area | 549 | 34.64 | 34.64 | 70.03
4 | Small town | 201 | 12.68 | 12.68 | 82.71
5 | Rural area | 274 | 17.29 | 17.29 | 100.00
<NA> | <NA> | 0 | 0.00 | <NA> | <NA>

Profile: 3 point Party ID (pid3) <numeric>
# total N=1585 valid N=1522 mean=2.21 sd=1.13

Value | Label | N | Raw % | Valid % | Cum. %
-----
1 | Democrat | 508 | 32.05 | 33.38 | 33.38
2 | Republican | 430 | 27.13 | 28.25 | 61.63
3 | Independent | 428 | 27.00 | 28.12 | 89.75
4 | Other | 66 | 4.16 | 4.34 | 94.09
5 | Not sure | 90 | 5.68 | 5.91 | 100.00
<NA> | <NA> | 63 | 3.97 | <NA> | <NA>

Profile: Ideology (ideo5) <numeric>
# total N=1585 valid N=1584 mean=3.34 sd=1.40

Value | Label | N | Raw % | Valid % | Cum. %

```

1	Very liberal	171 10.79 10.80 10.80
2	Liberal	253 15.96 15.97 26.77
3	Moderate	486 30.66 30.68 57.45
4	Conservative	342 21.58 21.59 79.04
5	Very conservative	198 12.49 12.50 91.54
6	Not sure	134 8.45 8.46 100.00
<NA>	<NA>	1 0.06 <NA> <NA>

Profile: Gender (gender) <numeric>
total N=1585 valid N=1585 mean=1.55 sd=0.50

Value	Label	N	Raw %	Valid %	Cum. %
1	Male	707	44.61	44.61	44.61
2	Female	878	55.39	55.39	100.00
<NA>	<NA>	0	0.00	<NA>	<NA>

Profile: Importance of religion (Pew version) (pew_religimp) <numeric>
total N=1585 valid N=1585 mean=2.24 sd=1.17

Value	Label	N	Raw %	Valid %	Cum. %
1	Very important	580	36.59	36.59	36.59
2	Somewhat important	416	26.25	26.25	62.84
3	Not too important	222	14.01	14.01	76.85
4	Not at all important	367	23.15	23.15	100.00
<NA>	<NA>	0	0.00	<NA>	<NA>

Follow what's going on in government and public affairs (follow) <numeric>
total N=1585 valid N=1585 mean=1.82 sd=0.97

Value	Label	N	Raw %	Valid %	Cum. %
1	Most of the time	773	48.77	48.77	48.77
2	Some of the time	460	29.02	29.02	77.79
3	Only now and then	217	13.69	13.69	91.48
4	Hardly at all	135	8.52	8.52	100.00
<NA>	<NA>	0	0.00	<NA>	<NA>

SA 2: WVS

Defining your research question, dataset, & variables

Find Quarto themes here.

Research Question

To what extent do socioeconomic variables and feelings about the honesty of politics explain how much confidence Australians have in their government?

Dataset Selected

World Values Survey — Australia

Variables Selected for Initial Study

Table 2: Table 1. Variables Chosen for Study

Purpose of Variable	Variable Name & Description	Level of Measurement
Dependent	Q71 - How much confidence do you have in the government?	ordinal
Independent	Q262 - age of the respondent X003R - recoded age X003R2 - recoded age, fewer categories Q112 - how much corruption is in the country? Q113 - belief that state authorities are corrupt Q199 - interest in politics	interval ordinal ordinal interval ordinal ordinal

Purpose of Variable	Variable Name & Description	Level of Measurement
	Q234 - does R think elections are honest	ordinal
	Q235 - want a strong leader with few checks?	ordinal
	Q260 - sex of R	nominal
	Q275R - recoded education	ordinal

Load packages and run function for selecting a country

```

library(tidyverse)
library(haven)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(ggblanket)

## Function for picking a country
pick_a_country <- function(country_to_filter) {
  read_sav("WVS_Cross-National_Wave_7_spss_v5_0.sav", user_na = FALSE) |>
    filter(B_COUNTRY_ALPHA == country_to_filter) |>
    drop_unused_value_labels()
}

```

Select a country to study, select variables, and save your new smaller dataset ([link to help](#))

Need help? [Go to chapter x in the webbook.](#)

```

# 1) Select a country from the master dataset. Replace ""3 letter code here" with the three letter code for the country you want to study.

australia <- pick_a_country("AUS")

# 2) Select variables in a country

australia_small <- australia |>

```

```

  select(Q71, Q112, Q113, Q199, Q234, Q235, Q260, Q275R, X003R, X003R2, W_WEIGHT)

# 3) Save original and smaller dataset

save(australia_small, file = "australia_small.RData")

```

Run variable frequencies on your new smaller dataset

Need help? [Go to chapter x in the webbook.](#)

```

australia_small |>
  select(-W_WEIGHT) |>
  frq()

```

Confidence: The Government (Q71) <numeric>
total N=1813 valid N=1796 mean=2.82 sd=0.75

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	A great deal	69	3.81	3.84	3.84
2	Quite a lot	493	27.19	27.45	31.29
3	Not very much	931	51.35	51.84	83.13
4	None at all	303	16.71	16.87	100.00
<NA>	<NA>	17	0.94	<NA>	<NA>

Perceptions of corruption in the country (Q112) <numeric>
total N=1813 valid N=1794 mean=6.65 sd=2.28

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	1 There is no corruption in my country	14	0.77	0.78	0.78
2		2	2.87	2.90	3.68
3		3	7.56	7.64	11.32
4		4	7.45	7.53	18.84
5		5	12.96	13.10	31.94
6		6	11.47	11.59	43.53
7		7	18.09	18.28	61.82
8		8	15.55	15.72	77.54
9		9	8.33	8.42	85.95
10	10 There is abundant corruption in my country	252	13.90	14.05	100.00

<NA> | <NA> | 19 | 1.05 | <NA> | <NA>

Involved in corruption: State authorities (Q113) <numeric>

total N=1813 valid N=1775 mean=2.19 sd=0.57

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	None of them	90	4.96	5.07	5.07
2	Few of them	1324	73.03	74.59	79.66
3	Most of them	302	16.66	17.01	96.68
4	All of them	59	3.25	3.32	100.00
<NA>	<NA>	38	2.10	<NA>	<NA>

Interest in politics (Q199) <numeric>

total N=1813 valid N=1802 mean=2.35 sd=0.89

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	Very interested	293	16.16	16.26	16.26
2	Somewhat interested	805	44.40	44.67	60.93
3	Not very interested	486	26.81	26.97	87.90
4	Not at all interested	218	12.02	12.10	100.00
<NA>	<NA>	11	0.61	<NA>	<NA>

Some people think that having honest elections makes a lot of difference in their lives; othe

total N=1813 valid N=1797 mean=1.25 sd=0.57

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	Very important	1450	79.98	80.69	80.69
2	Rather important	266	14.67	14.80	95.49
3	Not very important	61	3.36	3.39	98.89
4	Not at all important	20	1.10	1.11	100.00
<NA>	<NA>	16	0.88	<NA>	<NA>

Political system: Having a strong leader who does not have to bother with parliament and elec

total N=1813 valid N=1744 mean=3.16 sd=0.97

Value	Label	N	Raw %	Valid %	Cum. %
<hr/>					
1	Very good	124	6.84	7.11	7.11
2	Fairly good	328	18.09	18.81	25.92
3	Fairly Bad	430	23.72	24.66	50.57

4	Very bad	862 47.55 49.43 100.00
<NA>	<NA>	69 3.81 <NA> <NA>

Sex (Q260) <numeric>
total N=1813 valid N=1799 mean=1.61 sd=0.49

Value	Label	N	Raw %	Valid %	Cum. %
1	Male	704	38.83	39.13	39.13
2	Female	1095	60.40	60.87	100.00
<NA>	<NA>	14	0.77	<NA>	<NA>

Highest educational level: Respondent (recoded into 3 groups) (Q275R) <numeric>
total N=1813 valid N=1743 mean=2.36 sd=0.64

Value	Label	N	Raw %	Valid %	Cum. %
1	Lower	152	8.38	8.72	8.72
2	Middle	807	44.51	46.30	55.02
3	Higher	784	43.24	44.98	100.00
<NA>	<NA>	70	3.86	<NA>	<NA>

Age recoded (6 intervals) (X003R) <numeric>
total N=1813 valid N=1795 mean=4.33 sd=1.55

Value	Label	N	Raw %	Valid %	Cum. %
1	16-24	77	4.25	4.29	4.29
2	25-34	228	12.58	12.70	16.99
3	35-44	250	13.79	13.93	30.92
4	45-54	276	15.22	15.38	46.30
5	55-64	392	21.62	21.84	68.13
6	65 and more years	572	31.55	31.87	100.00
<NA>	<NA>	18	0.99	<NA>	<NA>

Age recoded (3 intervals) (X003R2) <numeric>
total N=1813 valid N=1795 mean=2.51 sd=0.67

Value	Label	N	Raw %	Valid %	Cum. %
1	16-29 years	179	9.87	9.97	9.97
2	30-49 years	513	28.30	28.58	38.55
3	50 and more years	1103	60.84	61.45	100.00

<NA> | <NA> | 18 | 0.99 | <NA> | <NA>

SA 2: ANES 2020

Defining your research question, dataset, & variables

Find Quarto themes here.

Research Question

To what extent can interest in political campaigns be explained by demographic and political variables?

Dataset Selected

2020 American National Election Study

Variables Selected for Initial Study

Table 3: Table 1. Variables Chosen for Study

Purpose of Variable	Variable Name & Description	Level of Measurement
Dependent	V201006 - How interested are you in following political campaigns?	ordinal
Independent	V201507x - age of the respondent V201600- gender of the respondent V201018- party of the respondent V201200 - ideology of the respondent V201549x - race of the respondent V203003 - region of the respondent	interval nominal nominal ordinal nominal nominal

Purpose of Variable	Variable Name & Description	Level of Measurement

Load packages

```
library(tidyverse)
library(haven)
library(naniar)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(ggblanket)
```

Load the correct dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("anes_timeseries_2020.RData")
```

Using the dataset you have loaded, select your variables, and save your new smaller dataset

Need help? [Go to chapter 4 in the webbook.](#)

```
smaller_anes_2020 <- anes_timeseries_2020 |>
  select(V200001, V200016b, V201006, V201507x, V201600, V201008, V201200, V201549x, V203000 |
    drop_unused_value_labels()

save(smaller_anes_2020, file = "smaller_anes_2020.RData")
```

Run variable frequencies on your new smaller dataset

Need help? [Go to chapter x in the webbook.](#)

```
smaller_anes_2020 |>
  select(-V200001, -V200016b) |>
  frq()
```

```
PRE: How interested in following campaigns (V201006) <numeric>
# total N=8280 valid N=8279 mean=1.61 sd=0.70
```

Value	Label	N	Raw %	Valid %	Cum. %
1	1. Very much interested	4320	52.17	52.18	52.18
2	2. Somewhat interested	2890	34.90	34.91	87.09
3	3. Not much interested	1069	12.91	12.91	100.00
<NA>	<NA>	1	0.01	<NA>	<NA>

```
PRE: SUMMARY: Respondent age (V201507x) <numeric>
# total N=8280 valid N=7932 mean=51.59 sd=17.21
```

Value	Label	N	Raw %	Valid %	Cum. %
18	18	35	0.42	0.44	0.44
19	19	52	0.63	0.66	1.10
20	20	46	0.56	0.58	1.68
21	21	51	0.62	0.64	2.32
22	22	57	0.69	0.72	3.04
23	23	75	0.91	0.95	3.98
24	24	92	1.11	1.16	5.14
25	25	104	1.26	1.31	6.45
26	26	108	1.30	1.36	7.82
27	27	132	1.59	1.66	9.48
28	28	120	1.45	1.51	10.99
29	29	131	1.58	1.65	12.64
30	30	142	1.71	1.79	14.44
31	31	109	1.32	1.37	15.81
32	32	117	1.41	1.48	17.28
33	33	123	1.49	1.55	18.84
34	34	142	1.71	1.79	20.63
35	35	152	1.84	1.92	22.54
36	36	144	1.74	1.82	24.36

37	37	149	1.80	1.88	26.24
38	38	152	1.84	1.92	28.15
39	39	151	1.82	1.90	30.06
40	40	139	1.68	1.75	31.81
41	41	151	1.82	1.90	33.71
42	42	113	1.36	1.42	35.14
43	43	116	1.40	1.46	36.60
44	44	111	1.34	1.40	38.00
45	45	116	1.40	1.46	39.46
46	46	119	1.44	1.50	40.96
47	47	106	1.28	1.34	42.30
48	48	105	1.27	1.32	43.62
49	49	123	1.49	1.55	45.17
50	50	154	1.86	1.94	47.11
51	51	128	1.55	1.61	48.73
52	52	111	1.34	1.40	50.13
53	53	117	1.41	1.48	51.60
54	54	123	1.49	1.55	53.15
55	55	140	1.69	1.77	54.92
56	56	127	1.53	1.60	56.52
57	57	136	1.64	1.71	58.23
58	58	145	1.75	1.83	60.06
59	59	154	1.86	1.94	62.00
60	60	168	2.03	2.12	64.12
61	61	139	1.68	1.75	65.87
62	62	154	1.86	1.94	67.81
63	63	156	1.88	1.97	69.78
64	64	155	1.87	1.95	71.73
65	65	180	2.17	2.27	74.00
66	66	170	2.05	2.14	76.15
67	67	142	1.71	1.79	77.94
68	68	140	1.69	1.77	79.70
69	69	158	1.91	1.99	81.69
70	70	126	1.52	1.59	83.28
71	71	147	1.78	1.85	85.14
72	72	145	1.75	1.83	86.96
73	73	147	1.78	1.85	88.82
74	74	94	1.14	1.19	90.00
75	75	93	1.12	1.17	91.17
76	76	89	1.07	1.12	92.30
77	77	81	0.98	1.02	93.32
78	78	64	0.77	0.81	94.13
79	79	63	0.76	0.79	94.92

80	80. Age 80 or older	403	4.87	5.08	100.00
<NA>	<NA>	348	4.20	<NA>	<NA>

PRE: What is your (R) sex? [revised] (V201600) <numeric>
total N=8280 valid N=8213 mean=1.54 sd=0.50

Value	Label	N	Raw %	Valid %	Cum. %
1	1. Male	3763	45.45	45.82	45.82
2	2. Female	4450	53.74	54.18	100.00
<NA>	<NA>	67	0.81	<NA>	<NA>

PRE: Where is R registered to vote (pre-election) (V201008) <numeric>
total N=8280 valid N=8270 mean=1.27 sd=0.61

Value	Label	N	Raw %	Valid %	Cum. %	
1	1. Registered at this address	6787	81.97	82.07	82.07	
2	2. Registered at a different address	765	9.24	9.25	91.32	
3	3. Not currently registered	718	8.67	8.68	100.00	
<NA>		<NA>	10	0.12	<NA>	<NA>

PRE: 7pt scale liberal-conservative self-placement (V201200) <numeric>
total N=8280 valid N=8257 mean=17.90 sd=33.50

Value	Label	N	Raw %	Valid %	Cum. %	
1	1. Extremely liberal	369	4.46	4.47	4.47	
2	2. Liberal	1210	14.61	14.65	19.12	
3	3. Slightly liberal	918	11.09	11.12	30.24	
4	4. Moderate; middle of the road	1818	21.96	22.02	52.26	
5	5. Slightly conservative	821	9.92	9.94	62.20	
6	6. Conservative	1492	18.02	18.07	80.27	
7	7. Extremely conservative	428	5.17	5.18	85.45	
99	99. Haven't thought much about this	1201	14.50	14.55	100.00	
<NA>		<NA>	23	0.28	<NA>	<NA>

PRE: SUMMARY: R self-identified race/ethnicity (V201549x) <numeric>
total N=8280 valid N=8178 mean=1.63 sd=1.24

Value	Label	N	Raw %
1	1. White, non-Hispanic	5963	72.02

2		2. Black, non-Hispanic	726	8.77
3		3. Hispanic	762	9.20
4	4. Asian or Native Hawaiian/other Pacific Islander, non-Hispanic alone	284	3.43	
5	5. Native American/Alaska Native or other race, non-Hispanic alone	172	2.08	
6		6. Multiple races, non-Hispanic	271	3.27
<NA>		<NA>	102	1.23

SAMPLE: Census region (V203003) <numeric>
total N=8280 valid N=8280 mean=2.64 sd=1.00

Value	Label	N	Raw %	Valid %	Cum. %
1	1. Northeast	1396	16.86	16.86	16.86
2	2. Midwest	1997	24.12	24.12	40.98
3	3. South	3081	37.21	37.21	78.19
4	4. West	1806	21.81	21.81	100.00
<NA>	<NA>	0	0.00	<NA>	<NA>

SA 3: Data Management Strategies

Find Quarto themes [here](#).

Discussion

- Select: I added a variable that tells us who the respondent voted for in the 2020 presidential election.
- Filter the variable `presvote20post` so the dataset now only includes people who voted for either Trump or Biden.
- Recode `Roe` so that there are two categories: favor and don't favor.
- Recode `urbanicity2` so that there are 3 categories: city, suburban, and rural.

Load Packages

```
library(tidyverse)
library(haven)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(ggblanket)
```

Load Your Datasets

Need help? [Go to chapter x in the webbook](#).

```
load("anes_pilot.RData")
load("anes_pilot_smaller.RData")
```

Select variables

Need help? [Go to chapter x in the webbook.](#)

```
another_anes_pilot_smaller <- anes_pilot |>
  select(roe, presvote20post, urbanicity2, pid3, ideo5, gender, pew_religimp, follow, case
```

Filter Your Dataset

Need help? [Go to chapter x in the webbook.](#)

```
another_anes_pilot_smaller <- another_anes_pilot_smaller |>
  filter(presvote20post < 3)
```

Recode a Variable

Need help? [Go to chapter 4 in the webbook.](#)

```
another_anes_pilot_smaller <- another_anes_pilot_smaller |>
  as_factor() |>
  mutate(roe_recode = fct_collapse(roe,
    "Favor" = c("Favor"),
    "Doesn't favor" = c("Oppose", "Neither favor nor oppose")))
```

Create Another New Variable

Need help? [Go to chapter 4 in the webbook.](#)

```
another_anes_pilot_smaller_2 <- another_anes_pilot_smaller |>
  as_factor() |>
  mutate(urbanicity2_recode = fct_collapse(urbanicity2,
    "City" = c("Big city", "Smaller city"),
    "Suburb" = c("Suburban area"),
    "Rural" = c("Small town", "Rural area")))
```

Run Frequencies for the New Variables You Created

Need help? [Go to chapter 4 in the webbook.](#)

```
another_anes_pilot_smaller_2 |>
  select(-caseid, -weight) |>
  frq()
```

```
Favor/oppose - overturn Roe v. Wade (roe) <categorical>
# total N=1135 valid N=1135 mean=1.75 sd=0.67
```

Value	N	Raw %	Valid %	Cum. %
Favor	434	38.24	38.24	38.24
Oppose	554	48.81	48.81	87.05
Neither favor nor oppose	147	12.95	12.95	100.00
<NA>	0	0.00	<NA>	<NA>

```
Profile: 2020 President Vote Post Election (presvote20post) <categorical>
# total N=1135 valid N=1135 mean=1.47 sd=0.50
```

Value	N	Raw %	Valid %	Cum. %
Joe Biden	607	53.48	53.48	53.48
Donald Trump	528	46.52	46.52	100.00
Jo Jorgensen	0	0.00	0.00	100.00
Howie Hawkins	0	0.00	0.00	100.00
Other	0	0.00	0.00	100.00
Did not vote for President	0	0.00	0.00	100.00
<NA>	0	0.00	<NA>	<NA>

```
Profile: Urban-rural status (urbanicity2) <categorical>
# total N=1135 valid N=1135 mean=2.96 sd=1.31
```

Value	N	Raw %	Valid %	Cum. %
Big city	213	18.77	18.77	18.77
Smaller city	165	14.54	14.54	33.30
Suburban area	409	36.04	36.04	69.34
Small town	151	13.30	13.30	82.64
Rural area	197	17.36	17.36	100.00
<NA>	0	0.00	<NA>	<NA>

Profile: 3 point Party ID (pid3) <categorical>
total N=1135 valid N=1098 mean=1.98 sd=0.96

Value		N	Raw %	Valid %	Cum. %
<hr/>					
Democrat		426	37.53	38.80	38.80
Republican		347	30.57	31.60	70.40
Independent		266	23.44	24.23	94.63
Other		42	3.70	3.83	98.45
Not sure		17	1.50	1.55	100.00
<NA>		37	3.26	<NA>	<NA>

Profile: Ideology (ideo5) <categorical>
total N=1135 valid N=1134 mean=3.18 sd=1.30

Value		N	Raw %	Valid %	Cum. %
<hr/>					
Very liberal		144	12.69	12.70	12.70
Liberal		195	17.18	17.20	29.89
Moderate		328	28.90	28.92	58.82
Conservative		271	23.88	23.90	82.72
Very conservative		168	14.80	14.81	97.53
Not sure		28	2.47	2.47	100.00
<NA>		1	0.09	<NA>	<NA>

Profile: Gender (gender) <categorical>
total N=1135 valid N=1135 mean=1.54 sd=0.50

Value		N	Raw %	Valid %	Cum. %
<hr/>					
Male		519	45.73	45.73	45.73
Female		616	54.27	54.27	100.00
<NA>		0	0.00	<NA>	<NA>

Profile: Importance of religion (Pew version) (pew_religimp) <categorical>
total N=1135 valid N=1135 mean=2.18 sd=1.17

Value		N	Raw %	Valid %	Cum. %
<hr/>					
Very important		450	39.65	39.65	39.65
Somewhat important		283	24.93	24.93	64.58
Not too important		154	13.57	13.57	78.15

Not at all important	248	21.85	21.85	100.00
<NA>	0	0.00	<NA>	<NA>

Follow what's going on in government and public affairs (follow) <categorical>
total N=1135 valid N=1135 mean=1.57 sd=0.80

Value	N	Raw %	Valid %	Cum. %
Most of the time	675	59.47	59.47	59.47
Some of the time	307	27.05	27.05	86.52
Only now and then	116	10.22	10.22	96.74
Hardly at all	37	3.26	3.26	100.00
<NA>	0	0.00	<NA>	<NA>

Favor/oppose - overturn Roe v. Wade (roe_recode) <categorical>
total N=1135 valid N=1135 mean=1.62 sd=0.49

Value	N	Raw %	Valid %	Cum. %
Favor	434	38.24	38.24	38.24
Doesn't favor	701	61.76	61.76	100.00
<NA>	0	0.00	<NA>	<NA>

Profile: Urban-rural status (urbanicity2_recode) <categorical>
total N=1135 valid N=1135 mean=1.97 sd=0.80

Value	N	Raw %	Valid %	Cum. %
City	378	33.30	33.30	33.30
Suburb	409	36.04	36.04	69.34
Rural	348	30.66	30.66	100.00
<NA>	0	0.00	<NA>	<NA>

Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)

```
save(another_anes_pilot_smaller_2, file = "another_anes_pilot_smaller_2.RData")
```

SA 4: Single Variable EDA

Find Quarto themes [here](#).

Discussion

Load Packages

```
library(tidyverse)
library(haven)
library(hrbrthemes)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(skimr)
library(ggblanket)
```

Load Your Dataset

Need help? [Go to chapter x in the webbook](#).

```
load("anes_pilot_small.RData")
```

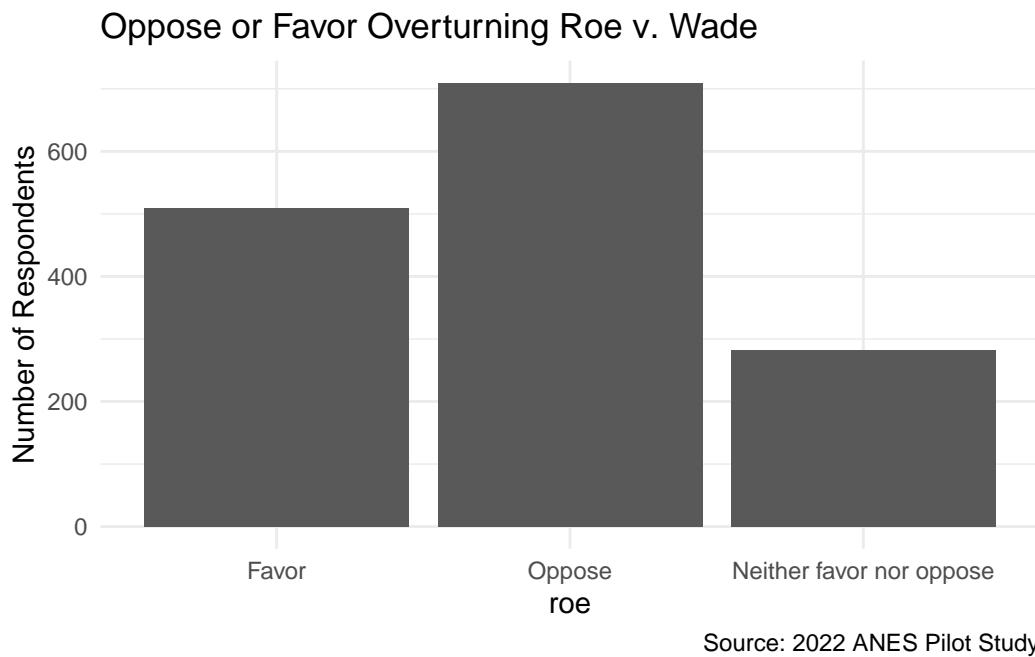
Manage your data as needed

Need help? [Go to chapter x in the webbook](#).

Graph 1

Need help? [Go to chapter 4 in the webbook.](#)

```
anes_pilot_small |>
  as_factor() |>
  drop_na(roe) |>
  ggplot(aes(x = roe, weight = weight)) +
  geom_bar() +
  ggtitle("Oppose or Favor Overturning Roe v. Wade") +
  labs(x = "roe", y = "Number of Respondents", caption = "Source: 2022 ANES Pilot Study")
  theme_minimal()
```



Graph 2

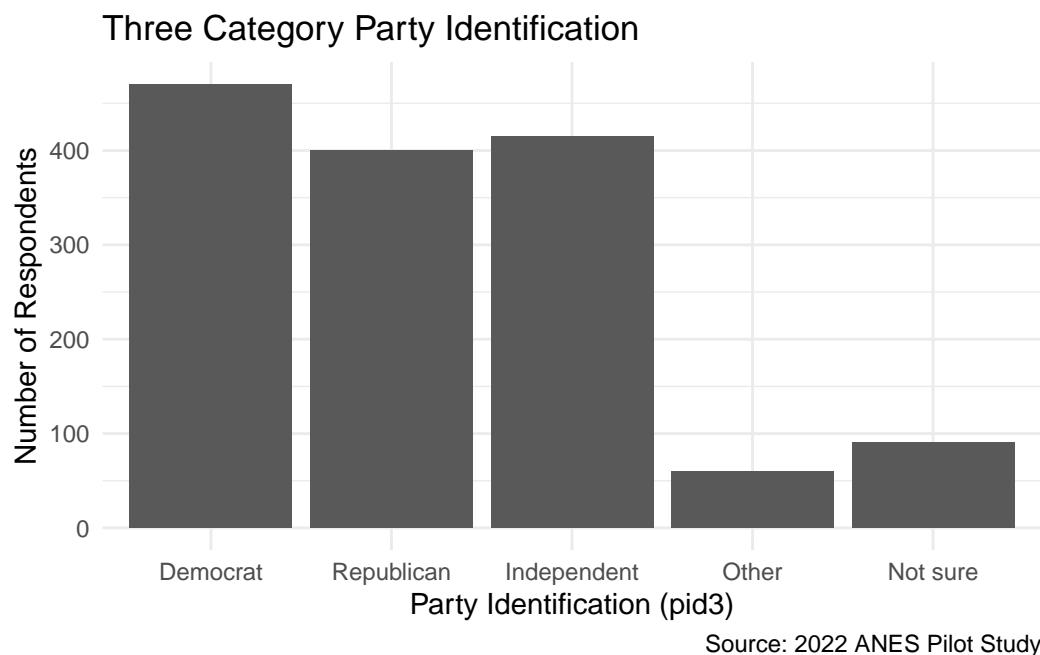
Need help? [Go to chapter 4 in the webbook.](#)

```
anes_pilot_small |>
  as_factor() |>
  drop_na(pid3) |>
  ggplot(aes(x = pid3, weight = weight)) +
```

```

geom_bar() +
ggtitle("Three Category Party Identification") +
labs(x = "Party Identification (pid3)", y = "Number of Respondents", caption = "Source:
theme_minimal()

```



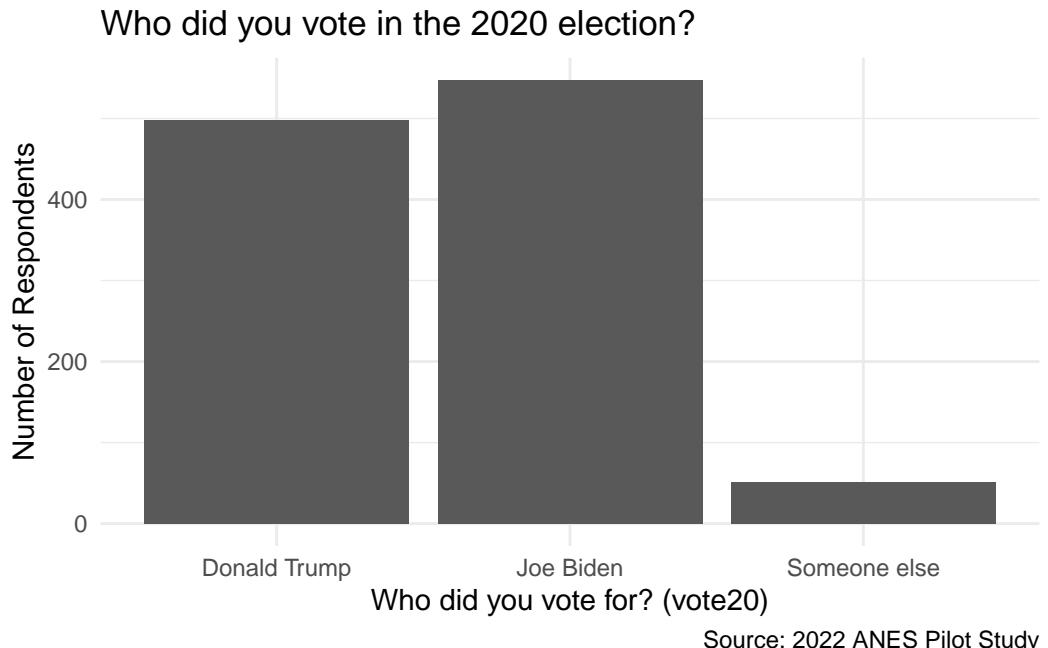
Graph 3 (add more if you want to!)

Need help? [Go to chapter 4 in the webbook](#).

```

anes_pilot_small |>
  as_factor() |>
  drop_na(vote20) |>
  ggplot(aes(x = vote20, weight = weight)) +
  geom_bar() +
  ggtitle("Who did you vote in the 2020 election?") +
  labs(x = "Who did you vote for? (vote20)", y = "Number of Respondents", caption = "Source:
  theme_minimal()

```



Run Univariate Statistics

Need help? [Go to chapter 4 in the webbook.](#)

```
anes_pilot_small |>
  as_factor() |>
 tbl_summary()
```

(1)
(2)

Table printed with `knitr::kable()`, not {gt}. Learn why at
<https://www.danielsgjoberg.com/gtsummary/articles/rmarkdown.html>
 To suppress this message, include `message = FALSE` in code chunk header.

```
anes_pilot_small |>
  as_factor() |>
  skim() |>
  yank("numeric")
```

Characteristic	N =
	1,585
Case ID	793 (397, 1,189)
Analysis weight	0.89 (0.65, 1.18)
Unknown	85
How would you rate people who forced their way into the U.S. Capitol on January 6, 2021?	15 (1, 50)
Unknown	2
How would you rate the U.S. Supreme Court?	51 (30, 74)
Compared to 2020, would you say the U.S. economy is much better, somewhat better, about the same, somewhat worse, or much worse?	
Much better	108 (6.8%)
Somewhat better	231 (15%)
About the same	273 (17%)
Somewhat worse	346 (22%)
Much worse	627 (40%)
Profile: Party Identification	
Strong Democrat	349 (23%)
Not very strong Democrat	156 (10%)
Lean Democrat	126 (8.3%)
Independent	245 (16%)
Lean Republican	150 (9.9%)
Not very strong Republican	150 (9.9%)
Strong Republican	281 (18%)

Characteristic	N =
	1,585
Not sure	62 (4.1%)
Unknown	66
Profile: 3 point Party ID	
Democrat	508 (33%)
Republican	430 (28%)
Independent	428 (28%)
Other	66 (4.3%)
Not sure	90 (5.9%)
Unknown	63
Profile: Ideology	
Very liberal	171 (11%)
Liberal	253 (16%)
Moderate	486 (31%)
Conservative	342 (22%)
Very conservative	198 (13%)
Not sure	134 (8.5%)
Unknown	1
Vote 2020 - Trump/Biden/Someone else	
Donald Trump	541 (45%)
Joe Biden	618 (51%)
Someone else	47 (3.9%)
Unknown	379
Follow what's going on in government and public affairs	
Most of the time	773 (49%)

Characteristic	N =
	1,585
Some of the time	460 (29%)
Only now and then	217 (14%)
Hardly at all	135 (8.5%)
Are you registered to vote, or not	
Yes, registered to vote at my current address	1,319 (83%)
Yes, registered to vote at a different address	70 (4.4%)
No, not registered	196 (12%)
Done the following things in the past 12 months – Attended a meeting to talk about political or social concern?	
Have done this in past 12 months	204 (13%)
Have not done this in the past 12 months	1,377 (87%)
Unknown	4
Done the following things in the past 12 months – Given money to any candidate running for public office, any political party, or any other group that supported or opposed candidates?	
Have done this in past 12 months	289 (18%)
Have not done this in the past 12 months	1,289 (82%)
Unknown	7
How important are each of the following issues in the country today? - Voting rights	
Extremely important	600 (38%)
Very important	364 (23%)
Moderately important	321 (20%)
Slightly important	177 (11%)
Not at all important	122 (7.7%)

Characteristic	N =
	1,585
Unknown	1
Favor/oppose - overturn Roe v. Wade	
Favor	526 (33%)
Oppose	764 (48%)
Neither favor nor oppose	295 (19%)
Profile: Gender	
Male	707 (45%)
Female	878 (55%)
Profile: Education	
No HS	75 (4.7%)
High school graduate	445 (28%)
Some college	357 (23%)
2-year	155 (9.8%)
4-year	363 (23%)
Post-grad	190 (12%)
Profile: Family income	
Less than \$10,000	112 (7.1%)
\$10,000 - \$19,999	120 (7.6%)
\$20,000 - \$29,999	157 (9.9%)
\$30,000 - \$39,999	124 (7.8%)
\$40,000 - \$49,999	107 (6.8%)
\$50,000 - \$59,999	121 (7.6%)

Characteristic	N =
	1,585
\$60,000 - \$69,999	108 (6.8%)
\$70,000 - \$79,999	119 (7.5%)
\$80,000 - \$99,999	123 (7.8%)
\$100,000 - \$119,999	81 (5.1%)
\$120,000 - \$149,999	75 (4.7%)
\$150,000 - \$199,999	82 (5.2%)
\$200,000 - \$249,999	37 (2.3%)
\$250,000 - \$349,999	27 (1.7%)
\$350,000 - \$499,999	10 (0.6%)
\$500,000 or more	7 (0.4%)
Prefer not to say	175 (11%)
Profile: Importance of religion (Pew version)	
Very important	580 (37%)
Somewhat important	416 (26%)
Not too important	222 (14%)
Not at all important	367 (23%)
Profile: Political Interest	
Most of the time	723 (46%)
Some of the time	429 (27%)
Only now and then	219 (14%)
Hardly at all	152 (9.6%)

Characteristic	N = 1,585
Don't know	60 (3.8%)
Unknown	2
Profile: Urban-rural status	
Big city	317 (20%)
Smaller city	244 (15%)
Suburban area	549 (35%)
Small town	201 (13%)
Rural area	274 (17%)

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
caseid	0	1.00	793.00	457.69	1.0	397.00	793.00	1189.00	1585.00	
weight	85	0.95	1.00	0.51	0.3	0.65	0.89	1.18	4.49	
jan6therm	2	1.00	27.50	29.46	0.0	1.00	15.00	50.00	100.00	
ftscotus	0	1.00	50.44	29.05	0.0	30.00	51.00	74.00	100.00	

Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)

SA 6: Two or more variable EDA

[Find Quarto themes here.](#)

Find the sample assignment here.

Discussion

I ran three graphs using my recoded *Roe v. Wade* variable.

- The first looked at whether support (or opposition) to overturning *Roe* varied by political party. Almost all Democrats opposed overturning *Roe*. A smaller majority of Republicans favored overturning *Roe*. The other groups were in the middle.
- The second graph looked at whether support (or opposition) to overturning *Roe* varied by where people lived. All groups had majorities of people who didn't favor overturning *Roe*. People lived in cities had the largest majorities, followed by suburban and rural Americans. The differences between the three groups were not as large as with party differences.
- The third graph examined whether support (or opposition) to overturning *Roe* varied by the importance of religion in people's lives. Majority of people for whom religion is very important favored overturning *Roe*. For the other groups a majority didn't favor overturning *Roe*. The percent of opposition went up as religion was less important in people's lives.
- None of the results I found surprised me.

Load Packages

```
library(tidyverse)
library(haven)
library(hrbrthemes)
library(survey)
library(srvyr)
library(labelled)
```

```
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(skimr)
library(ggblanket)
```

Load Your Dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("another_anes_pilot_smaller_2.RData")
```

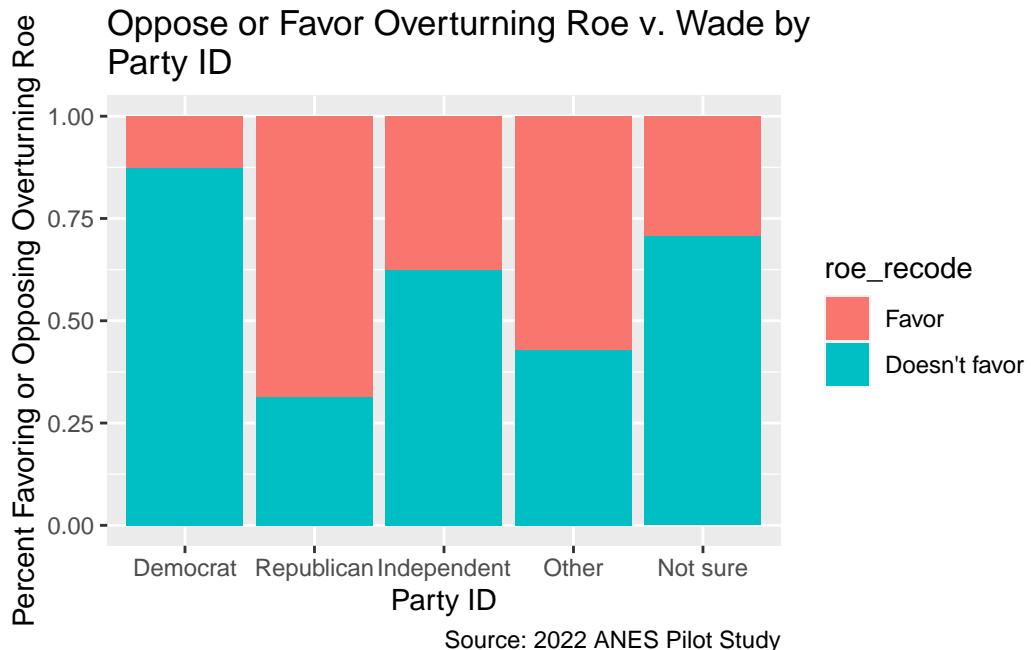
Manage your data as needed

Need help? [Go to chapter x in the webbook.](#)

Graph & Cross-tab 1

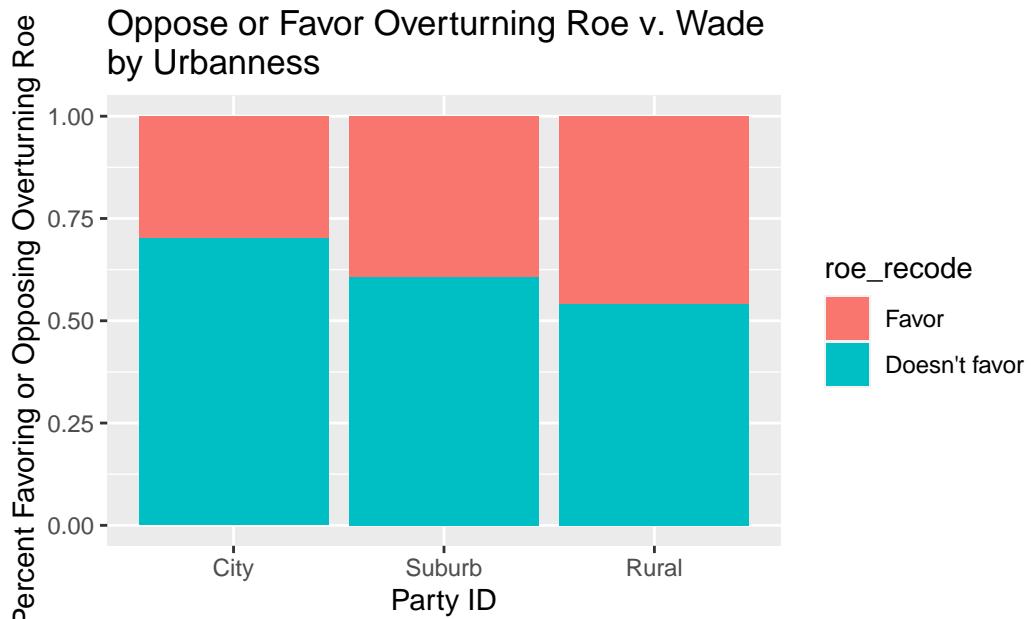
Need help? [Go to chapter 4 in the webbook.](#)

```
another_anes_pilot_smaller_2 |>
  as_factor() |>
  drop_na(pid3, roe_recode) |>
  ggplot(aes(x = pid3, fill = roe_recode)) +
  geom_bar(position= "fill") +
  ggtitle("Oppose or Favor Overturning Roe v. Wade by \nParty ID") +
  labs(x = "Party ID", y = "Percent Favoring or Opposing Overturning Roe", caption = "Sourc
```



Graph & Cross-tab 2

```
another_anes_pilot_smaller_2 |>
  as_factor() |>
  drop_na(urbanicity2_recode, roe_recode) |>
  ggplot(aes(x = urbanicity2_recode, fill = roe_recode)) +
  geom_bar(position= "fill") +
  ggtitle("Oppose or Favor Overturning Roe v. Wade \nby Urbanness") +
  labs(x = "Party ID", y = "Percent Favoring or Opposing Overturning Roe", caption = "Sour
```



```
tab_xtab(another_anes_pilot_smaller_2$roe_recode, another_anes_pilot_smaller_2$urbanicity2)
```

Favor/oppose -
overturn Roe v. Wade

Profile: Urban-rural
status

Total

City

Suburb

Rural

Favor

106
31.7 %

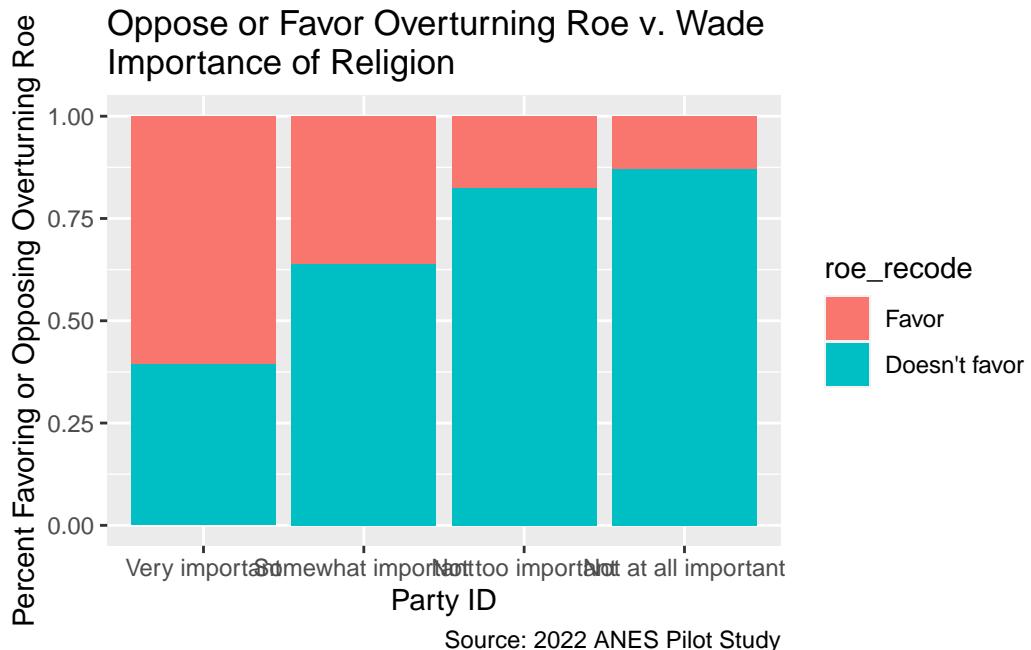
146
41 %

145
47.9 %

397	
40 %	
Doesn't favor	
228	
68.3 %	
210	
59 %	
158	
52.1 %	
596	
60 %	
Total	
334	
100 %	
356	
100 %	
303	
100 %	
993	
100 %	
$\chi^2 = 17.447 \cdot df=2 \cdot$ Cramer's V=0.133 · p=0.000	

Graph & Cross-tab 3 (add more if you want to!)

```
another_anes_pilot_smaller_2 |>
  as_factor() |>
  drop_na(pew_religimp, roe_recode) |>
  ggplot(aes(x = pew_religimp, fill = roe_recode)) +
  geom_bar(position= "fill") +
  ggtitle("Oppose or Favor Overturning Roe v. Wade \nImportance of Religion") +
  labs(x = "Party ID", y = "Percent Favoring or Opposing Overturning Roe", caption = "Sourc
```



Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)

SA 7: Hypothesis Testing

[Find Quarto themes here.](#)

Find the sample assignment here.

Discussion

Load Packages

```
library(tidyverse)
library(haven)
library(hrbrthemes)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(skimr)
library(ggblanket)
```

Load Your Dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("another_anes_pilot_smaller_2.RData")
```

Manage your data as needed

Need help? [Go to chapter x in the webbook.](#)

Hypothesis Test 1

Need help? [Go to chapter 4 in the webbook.](#)

```
tab_xtab(another_anes_pilot_smaller_2$roe_recode, another_anes_pilot_smaller_2$pid3, show=
```

Favor/oppose -

overturn Roe v. Wade

Profile: 3 point

Party ID

Total

Democrat

Republican

Independent

Other

Not sure

Favor

58

15.8 %

213

69.8 %

85

36.5 %

22

61.1 %

6

37.5 %

384

40.1 %

Doesn't favor

309

84.2 %

92

30.2 %

148	
63.5 %	
14	
38.9 %	
10	
62.5 %	
573	
59.9 %	
Total	
367	
100 %	
305	
100 %	
233	
100 %	
36	
100 %	
16	
100 %	
957	
100 %	

$\chi^2 = 210.359 \cdot df=4 \cdot$ Cramer's V=0.469 · p=0.000

Hypothesis Test 2

```
tab_xtab(another_anes_pilot_smaller_2$roe_recode, another_anes_pilot_smaller_2$urbanicity2
```

Favor/oppose -
overturn Roe v. Wade

Profile: Urban-rural
status

Total

City

Suburb

Rural	
Favor	
106	
31.7 %	
146	
41 %	
145	
47.9 %	
397	
40 %	
Doesn't favor	
228	
68.3 %	
210	
59 %	
158	
52.1 %	
596	
60 %	
Total	
334	
100 %	
356	
100 %	
303	
100 %	
993	
100 %	
$\chi^2=17.447$	· df=2 · Cramer's V=0.133
	· p=0.000

Hypothesis Test 3 (add more if needed!)

```
tab_xtab(another_anes_pilot_smaller_2$roe_recode, another_anes_pilot_smaller_2$pew_religim
```

Favor/oppose -
overturn Roe v. Wade

Profile: Importance
of religion (Pew
version)

Total

Very important

Somewhat important

Not too important

Not at all important

Favor

247
62.5 %

92
37.1 %

28
20.3 %

30
14.2 %

397
40 %

Doesn't favor

148
37.5 %

156
62.9 %

110
79.7 %

182
85.8 %
596
60 %
Total
395
100 %
248
100 %
138
100 %
212
100 %
993
100 %
$\chi^2 = 165.814 \cdot df=3 \cdot$ Cramer's V=0.409 · p=0.000

Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)

SA 8: Logistic Regression Analysis

[Find Quarto themes here.](#)

Find the sample assignment here.

Discussion

Load Packages

```
library(tidyverse)
library(haven)
library(hrbrthemes)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(skimr)
library(ggblanket)
```

Load Your Dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("anes_pilot_small.RData")
```

Manage your data as needed

Need help? [Go to chapter x in the webbook.](#)

Final Regression Model

Need help? [Go to chapter 4 in the webbook.](#)

```
opinion_of_scotus <- lm(ftscotus ~ as_factor(ideo5) + as_factor(urbanicity2) + as_factor(pew_religimp), data = anes_pilot_small)
summary(opinion_of_scotus)
```

Call:

```
lm(formula = ftscotus ~ as_factor(ideo5) + as_factor(urbanicity2) +
  as_factor(pew_religimp), data = anes_pilot_small)
```

Residuals:

Min	1Q	Median	3Q	Max
-71.159	-16.889	1.078	17.119	61.433

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	43.121	2.498	17.264
as_factor(ideo5)Liberal	3.173	2.490	1.274
as_factor(ideo5)Moderate	16.291	2.273	7.167
as_factor(ideo5)Conservative	30.636	2.473	12.386
as_factor(ideo5)Very conservative	32.782	2.781	11.786
as_factor(ideo5)Not sure	12.380	2.893	4.280
as_factor(urbanicity2)Smaller city	-4.743	2.133	-2.223
as_factor(urbanicity2)Suburban area	-4.872	1.773	-2.747
as_factor(urbanicity2)Small town	-2.652	2.256	-1.175
as_factor(urbanicity2)Rural area	-2.952	2.096	-1.409
as_factor(pew_religimp)Somewhat important	-5.885	1.644	-3.579
as_factor(pew_religimp)Not too important	-10.765	2.033	-5.295
as_factor(pew_religimp)Not at all important	-15.682	1.811	-8.660
	Pr(> t)		
(Intercept)	< 2e-16 ***		
as_factor(ideo5)Liberal	0.202734		
as_factor(ideo5)Moderate	1.17e-12 ***		
as_factor(ideo5)Conservative	< 2e-16 ***		
as_factor(ideo5)Very conservative	< 2e-16 ***		
as_factor(ideo5)Not sure	1.98e-05 ***		
as_factor(urbanicity2)Smaller city	0.026327 *		
as_factor(urbanicity2)Suburban area	0.006080 **		
as_factor(urbanicity2)Small town	0.240005		
as_factor(urbanicity2)Rural area	0.159091		

```

as_factor(pew_religimp)Somewhat important  0.000355 ***
as_factor(pew_religimp)Not too important   1.36e-07 ***
as_factor(pew_religimp)Not at all important < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 24.9 on 1571 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared: 0.2712, Adjusted R-squared: 0.2657
F-statistic: 48.73 on 12 and 1571 DF, p-value: < 2.2e-16

```

tbl_regression(opinion_of_scotus) |>
add_glance_source_note()

```

Characteristic	Beta	95% CI	p-value
Profile: Ideology	—	—	
Very liberal	—	—	
Liberal	3.2	-1.7, 8.1	0.2
Moderate	16	12, 21	<0.001
Conservative	31	26, 35	<0.001
Very conservative	33	27, 38	<0.001
Not sure	12	6.7, 18	<0.001
Profile: Urban-rural status	—	—	
Big city	—	—	
Smaller city	-4.7	-8.9, -0.56	0.026
Suburban area	-4.9	-8.3, -1.4	0.006
Small town	-2.7	-7.1, 1.8	0.2
Rural area	-3.0	-7.1, 1.2	0.2
Profile: Importance of religion (Pew version)	—	—	
Very important	—	—	
Somewhat important	-5.9	-9.1, -2.7	<0.001
Not too important	-11	-15, -6.8	<0.001
Not at all important	-16	-19, -12	<0.001

```

tab_model(opinion_of_scotus)

```

How would you rate the U S Supreme Court?			
Predictors	Estimates	CI	p

(Intercept)	43.12	38.22 – 48.02	<0.001
as_factor(ideo5)Liberal	3.17	-1.71 – 8.06	0.203
as_factor(ideo5)Moderate	16.29	11.83 – 20.75	<0.001
as_factor(ideo5)Conservative	30.64	25.78 – 35.49	<0.001
as_factor(ideo5)Very conservative	32.78	27.33 – 38.24	<0.001
as_factor(ideo5)Not sure	12.38	6.71 – 18.05	<0.001
as_factor(urbanicity2)Smaller city	-4.74	-8.93 – -0.56	0.026
as_factor(urbanicity2)Suburban-4.87 area		-8.35 – -1.39	0.006
as_factor(urbanicity2)Small town	-2.65	-7.08 – 1.77	0.240
as_factor(urbanicity2)Rural area	-2.95	-7.06 – 1.16	0.159
as_factor(pew_religimp)Somewhat important	5.88	-9.11 – -2.66	<0.001
as_factor(pew_religimp)Not too important	-10.77	-14.75 – -6.78	<0.001
as_factor(pew_religimp)Not at all important	-15.68	-19.23 – -12.13	<0.001
Observations	1584		
R ² / R ² adjusted	0.271 / 0.266		

Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)

SA 8: Linear Regression Analysis

[Find Quarto themes here.](#)

Find the sample assignment here.

Discussion

Load Packages

```
library(tidyverse)
library(haven)
library(hrbrthemes)
library(survey)
library(srvyr)
library(labelled)
library(sjmisc)
library(sjPlot)
library(gmodels)
library(gtsummary)
library(skimr)
library(ggblanket)
```

Load Your Dataset

Need help? [Go to chapter x in the webbook.](#)

```
load("another_anes_pilot_smaller_2.RData")
```

Manage your data as needed

Need help? [Go to chapter x in the webbook.](#)

Final Regression Model

Need help? [Go to chapter 4 in the webbook.](#)

```
roe_glm <- glm(roe_recode ~ ideo5 + urbanicity2_recode + pew_religimp, data = another_anes
summary(roe_glm)
```

Call:

```
glm(formula = roe_recode ~ ideo5 + urbanicity2_recode + pew_religimp,
family = "binomial", data = another_anes_pilot_smaller_2,
weights = another_anes_pilot_smaller_2$weight)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)		
(Intercept)	1.41299	0.34234	4.127	3.67e-05 ***		
ideo5Liberal	0.09176	0.41695	0.220	0.826		
ideo5Moderate	-0.98755	0.34309	-2.878	0.004 **		
ideo5Conservative	-2.86078	0.35038	-8.165	3.22e-16 ***		
ideo5Very conservative	-3.48971	0.39194	-8.904	< 2e-16 ***		
ideo5Not sure	0.25586	0.74398	0.344	0.731		
urbanicity2_recodeSuburb	-0.06706	0.20903	-0.321	0.748		
urbanicity2_recodeRural	-0.16370	0.21899	-0.747	0.455		
pew_religimpSomewhat important	0.94584	0.20458	4.623	3.78e-06 ***		
pew_religimpNot too important	1.54525	0.27888	5.541	3.01e-08 ***		
pew_religimpNot at all important	1.47602	0.26265	5.620	1.91e-08 ***		

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1334.70 on 1055 degrees of freedom
Residual deviance: 873.07 on 1045 degrees of freedom

(79 observations deleted due to missingness)

AIC: 976.48

Number of Fisher Scoring iterations: 5

```
tbl_regression(roe_glm, exponentiate = TRUE) |>
  add_glance_source_note(
    include = c(AIC))
```

Characteristic	OR	95% CI	p-value
Profile: Ideology	—	—	
Very liberal	—	—	
Liberal	1.10	0.48, 2.48	0.8
Moderate	0.37	0.18, 0.71	0.004
Conservative	0.06	0.03, 0.11	<0.001
Very conservative	0.03	0.01, 0.06	<0.001
Not sure	1.29	0.34, 6.99	0.7
Profile: Urban-rural status	—	—	
City	—	—	
Suburb	0.94	0.62, 1.41	0.7
Rural	0.85	0.55, 1.31	0.5
Profile: Importance of religion (Pew version)	—	—	
Very important	—	—	
Somewhat important	2.57	1.73, 3.86	<0.001
Not too important	4.69	2.74, 8.21	<0.001
Not at all important	4.38	2.64, 7.40	<0.001

```
tab_model(roe_glm)
```

Favor/oppose-overturn Roe v Wade			
Predictors	Odds Ratios	CI	p
(Intercept)	4.11	2.17 – 8.37	<0.001
Profile:Ideology:	1.10	0.48 – 2.48	0.826
Liberal	0.37	0.18 – 0.71	0.004
Moderate	0.06	0.03 – 0.11	<0.001
Conservative	0.03	0.01 – 0.06	<0.001
Very conservative	1.29	0.34 – 6.99	0.731
Profile:Ideology: Not sure	0.94	0.62 – 1.41	0.748
Profile:Urban-rural status: Suburb	0.85	0.55 – 1.31	0.455
Profile:Urban-rural status: Rural	—	—	

Profile:Importance of religion(Pew version):	2.57	1.73 – 3.86	<0.001
Somewhat important			
Profile:Importance of religion(Pew version):	4.69	2.74 – 8.21	<0.001
Not too important			
Profile:Importance of religion(Pew version):	4.38	2.64 – 7.40	<0.001
Not at all important			
Observations	1056		
R ² Tjur	0.012		

Save your updated dataset?

Need help? [Go to chapter 4 in the webbook.](#)