Quantitative Sociological Analysis

Descriptive Statistics

Exercise 3

February 11-13, 2025

Summarizing descriptive statistics: Netflix

- Let's start to make sense of our Netflix survey data
  - and work toward making a descriptive statistics table

- First, we need to consider each variable's level of measurement
  - so we know how to appropriately summarize the data

Exercise 3

Let's work in our groups. Review the survey and try to determine how responses to each respective question should be coded into a variable so we can make sense of these data.

Assign a scribe to take notes. Scratch paper will be fine, or maybe download this PPT and use that. You won't turn this part in, but we will share our thoughts as a class afterward.

age

- date of birth is a standard survey question because it is more reliable and valid than asking a respondent to report their age
  - age can be computed based on survey date – birth date
    - this also allows for age to be measured in days when precision is necessary

- age will be recorded in years
  - level of measurement: interval-ratio

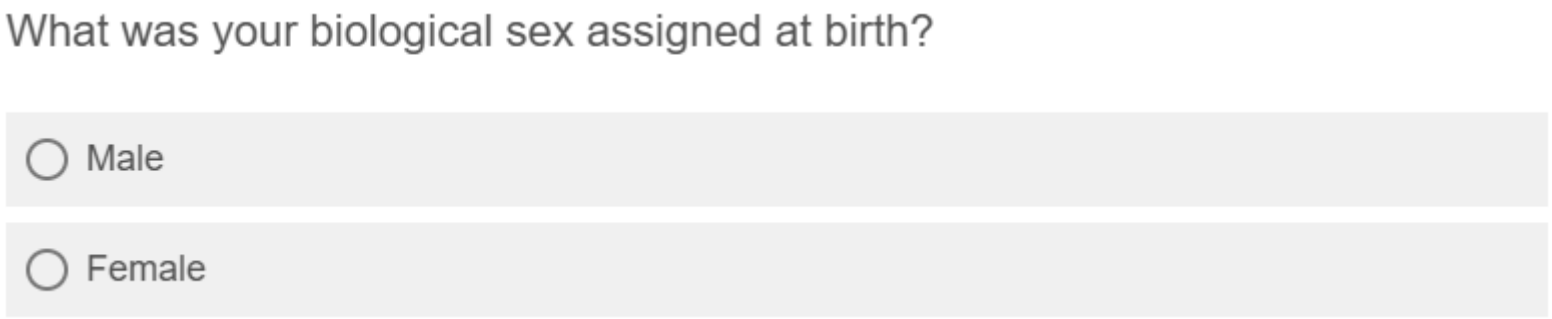
subjective age

- response options range from 0 to 100 years old
  - we don’t need to convert this variable

- subjective age will be recorded in years
  - level of measurement: _____

Fill in the blank here, and other blanks
as we move forward in this exercise

sex

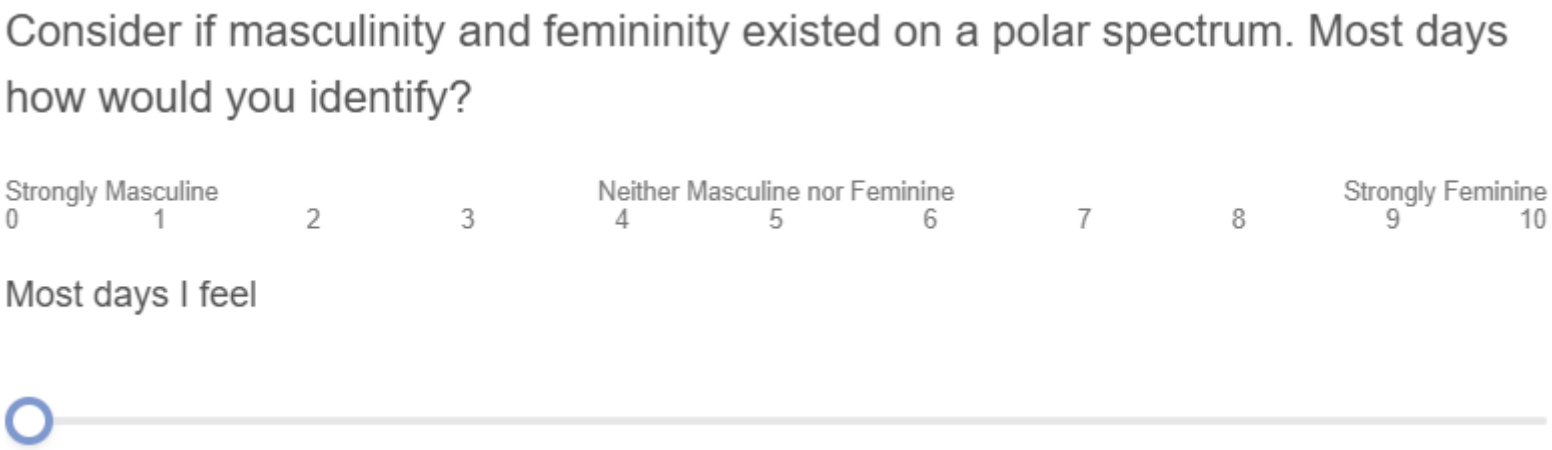


- includes two mutually exclusive response options
  - example of a binary indicator variable
    - Why should this type of variable be coded 0 1, rather than 1 2?

Answer the question, and other questions as we move forward in this exercise

- sex will be coded as 0 = male, 1 = female
  - Why will we name this variable female?
    - Hint: mean female = 0.73. Thus, what proportion are male?
  - level of measurement: _____

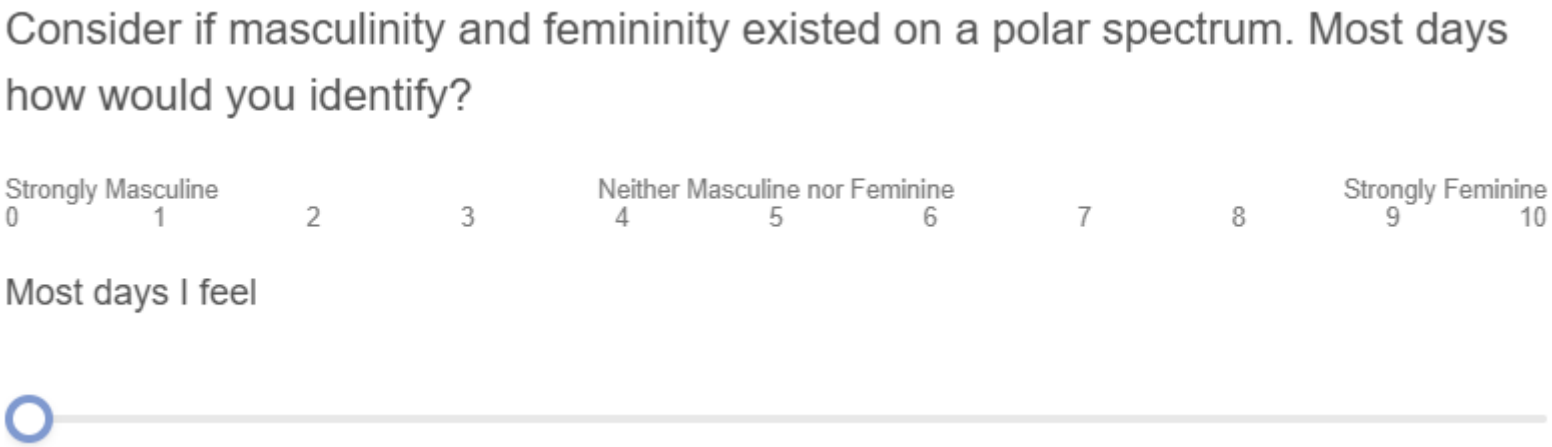
gender



- response options range from 0 to 10 on a spectrum
  - this provides several options for how to treat this measure…

gender

option (a)

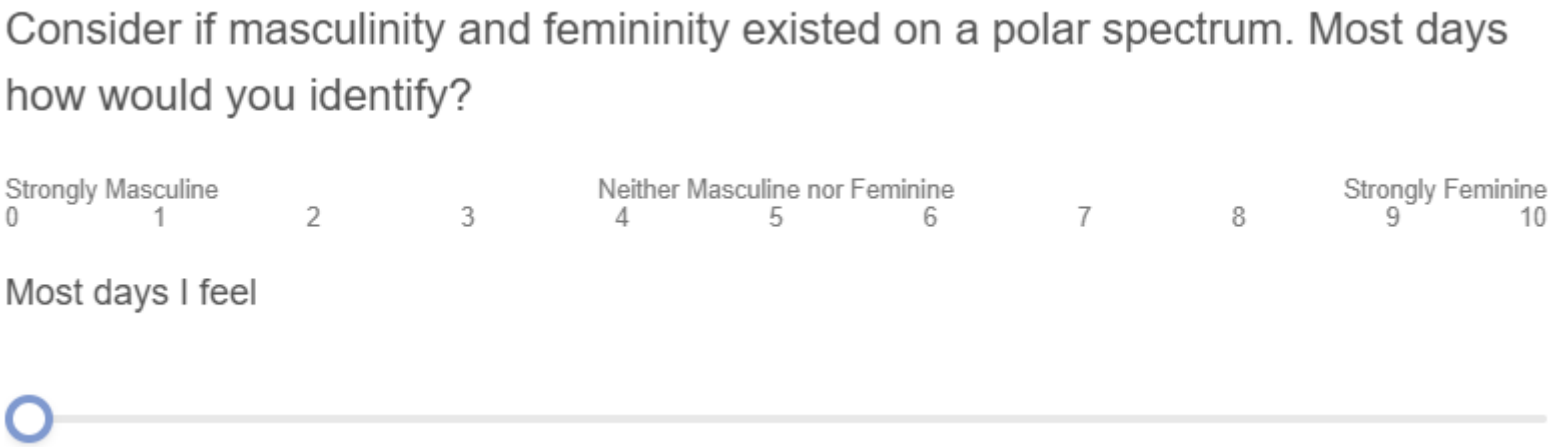


- we will retain this measurement for our variable
  - Thus, higher scores reflect _____, whereas lower scores reflect _____, and score in the middle of this spectrum reflects _____.

- Level of measurement: response options can be rank ordered, so we could treat as ordinal
  - However, ordinal measures with many categories are often treated like interval-ratio
    - We will treat gender like an interval-ratio variable

gender

option (b)

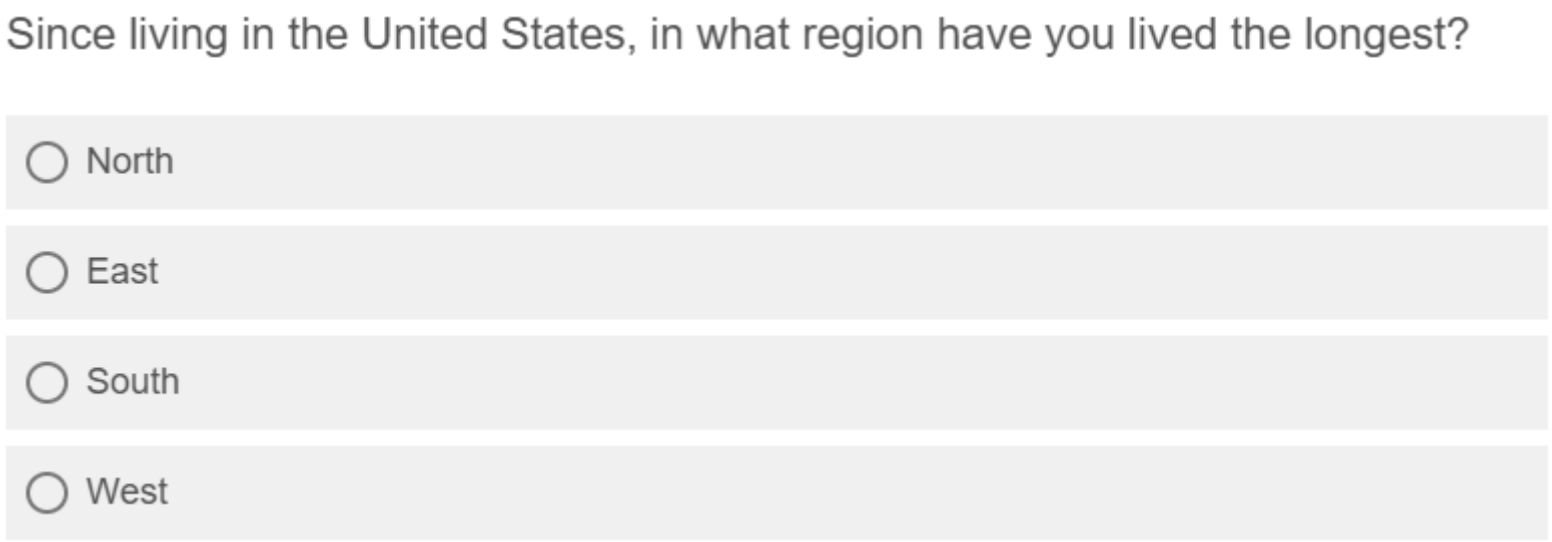


- if we wanted to treat gender like an ordinal measure it could be helpful to collapse some response options so there were fewer categories

- What might that type of gender variable look like?
  - Note: many possible “correct” answers
    - Recall, a variable should best reflect its conceptualization

rooted in perspective and theory

region



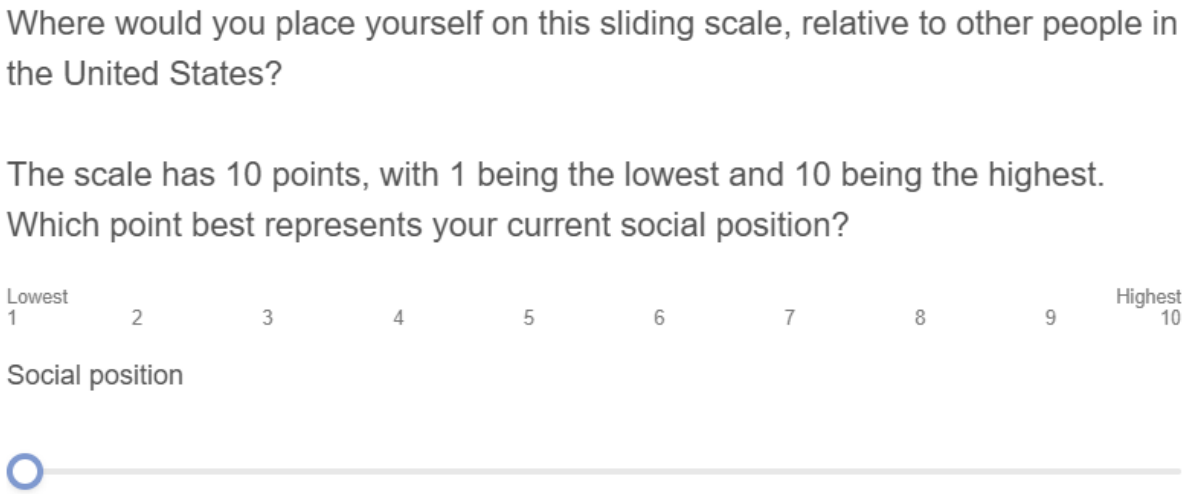
- How should region be coded?
  - level of measurement: _____
    - Hint: the numbers don't really matter, so long as they are mutually exclusive
      - but ideal for coding to be somewhat intuitive

rurality

- will be coded as 1=urban, 2=suburban, 3=rural
  - level of measurement: _____

- Why would we want to name this variable urbanicity if it were coded as
  - 1=rural, 2=suburban, 3=urban?
    - Hint: like with the region variable, the numbers don't really matter but ideal to be intuitive
      - Why might naming this variable suburban as 1=rural, 2=urban, 3=suburban not be very intuitive?

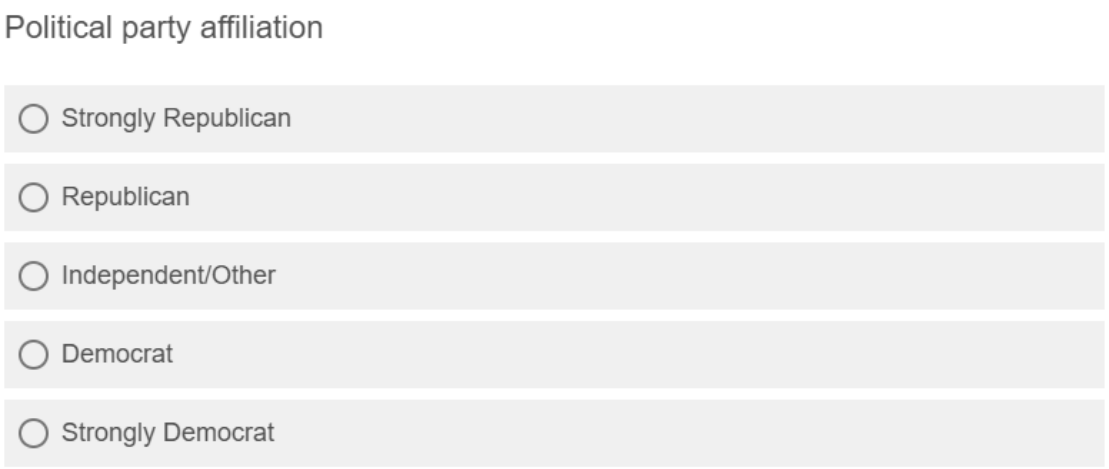
subjective socioeconomic status



- response options range from 1 (lowest) to 10 (highest), with equal intervals
  - because of so many response categories we will treat this variable as if it were…
    - level of measurement: _____

- What if we collapsed categories to create a ses variable like this…
  - 1-4 (1)=lower class, 5-7 (2)=middle class, 8-10 (3)=upper class
    - level of measurement: _____

political party affiliation



- How would you code this variable, and what would you name it?
  - level of measurement: _____
    - Note: many possible “correct” answers

composite variable

- a new variable created by combining other variables
  - often to reflect a broader construct
    - commonly referred to as a scale or an index

- sometimes based on well-established guidelines
  - like with many psychometric scales (e.g., anxiety, depression, attention)

anxiety

How often have you experienced the following during the past two weeks?

	Not at all	Several days	More than half the days	Nearly every day
Feeling nervous, anxious, or on edge	○	○	○	○
Not being able to stop or control worrying	○	○	○	○
Worrying too much about different things	○	○	○	○
Trouble relaxing	○	○	○	○
Being so restless that it is hard to sit still	○	○	○	○
Becoming easily annoyed or irritable	○	○	○	○
Feeling afraid as if something awful might happen	○	○	○	○

- based on the GAD-7 guidelines…
- how should this variable be coded?
  - level of measurement: _____
- If we were to use the clinical threshold, which is ≥10
  - how would this binary indicator variable be coded?
    - level of measurement: _____

depression

How often have you experienced the following during the past two weeks?

	Not at all	Several days	More than half the days	Nearly every day
Little interest or pleasure in doing things	○	○	○	○
Feeling down, depressed, or hopeless	○	○	○	○
Trouble falling or staying asleep, or sleeping too much	○	○	○	○
Feeling tired or having little energy	○	○	○	○
Poor appetite or overeating	○	○	○	○
Feeling bad about yourself — or that you are a failure or have let yourself or your family down	○	○	○	○
Trouble concentrating on things, such as reading the newspaper or watching television	○	○	○	○
Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	○	○	○	○
Thoughts that you would be better off dead, or of hurting yourself in some way	○	○	○	○

- based on the PHQ-9 guidelines…
- how should this variable be coded?
  - too complex, just use box at bottom
    - level of measurement: _____
- For simplicity let's use a clinical threshold of ≥10 for a binary indicator

attention


How often do you experience the following?

	Never	Rarely	Sometimes	Often	Very Often
make careless mistakes when you have to work on a boring or difficult project	○	○	○	○	○
have trouble keeping your attention focused when you are doing boring or repetitive work	○	○	○	○	○
have difficulty concentrating on what people say to you, even when they are speaking to you directly	○	○	○	○	○
have difficulty organizing tasks and activities	○	○	○	○	○
avoid, dislike, or are reluctant to engage in tasks that require sustained mental effort	○	○	○	○	○
forget to complete tasks	○	○	○	○	○

- based on the ASRS guidelines…
  - Part A only
- how should this variable be coded?
  - level of measurement: _____

macro sociohistorical experience

- due to my poor operationalization, it’s not worth wrapping our heads around how to construct this variable
  - unless you have any good ideas

preference for comedy

- we will construct this variable based on the rank ordered placement of comedy
  - 1=comedy is most favorite, 2=comedy is second favorite, 3=comedy is third favorite, 4=comedy is fourth favorite, 5=comedy is least favorite
    - level of measurement: _____

Note how this approach regrettably disregards potentially useful information – the rank order of the other four specific genres is likely useful for predicting Netflix consumer behavior. However, how complex might it be to include this information? Could this information be reflected in just one variable?

genre rankings

Not concerned with this for now but may see something like this later.

- Permutations: all possible unique arrangements when order matters

$$P(n,n) = \frac{n!}{(n-n)!} = n!$$

- There are 5 genres
  - 5! = 120

What did we all come up with?

Summarizing descriptive statistics: Netflix

Descriptive Statistics Table: Netflix Survey SOC303 Spring 2025 (N = 22)

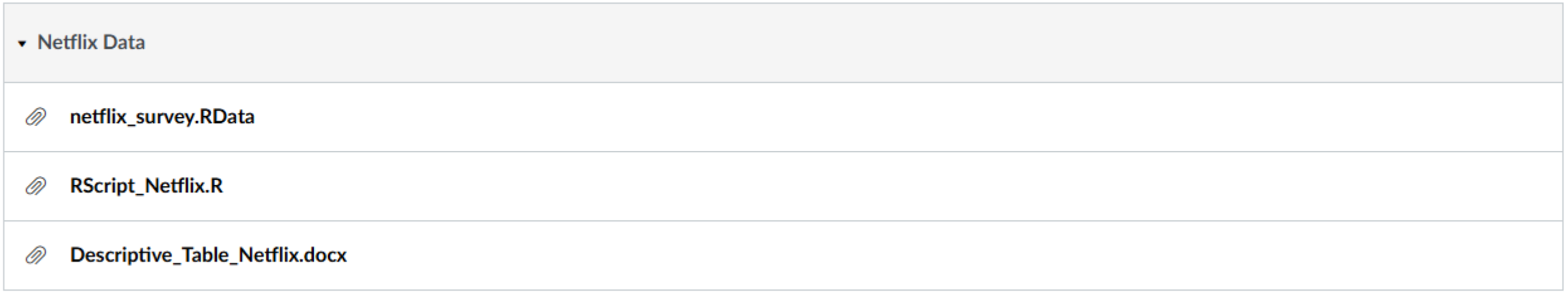
Variable	Mean (SD)	Median	Min.	Max.	Level of Measurement
Comedy Preference [a]			1	5	ordinal
Most Favorite	0.18				
Second Favorite	0.27				
Third Favorite	0.27				
Fourth Favorite	0.23				
Least Favorite	0.05				
Comedy Preference [b]	2.68	3.00	1	5	interval-ratio
Age					
Subjective Age					
Female					
Gender					
Region					
North					
East					
South					
West					
Rurality					
Urban					
Suburban					
Rural					
Subjective SES					
Political Party Affiliation					
Strongly Republican					
Republican					
Independent/Other					
Democrat					
Strongly Democrat					
Anxiety					
Minimal					
Mild					
Moderate					
Severe					
Anxiety Binary					
Depression					
Minimal					
Mild					
Moderate					
Moderately Severe					
Depression Binary					
Attention Binary					
Macro Experience?					

Notes: [a,b] we may want to treat Comedy Preference like an ordinal [a] or an interval-ratio [b] variable, so let's consider descriptive statistics for both levels of measurement.

- Now we are ready to start making a descriptive statistics table for our Netflix survey data
  - I made some executive decisions on variable construction, but we can revisit later if you want
- I started us off by providing a skeleton and summarizing our DV
  - Comedy Preference
- Let's see how this goes working in our groups
  - then I'll design Assignment 1 accordingly

Maybe push back due date because of snow day

Summarizing descriptive statistics: Netflix



- everything you need is in our Netflix Data module on Canvas
  - Try a variable or two as a team, then divide the remaining variables amongst your group as individuals or pairs/trios.
    - You're a team. Help each other when stuck. If all stuck, ask me.

It may be helpful to keep a working draft of your own RScript and descriptive statistics table. Then share programming and results with one another, maybe even check each others work. Make sure to provide kind and constructive feedback. Each team should have a completed descriptive statistics table when we are finished.