Measurement and Plotting in R

Today's Goals

Review:

- interpretation/usefulness of different plots
- plotting syntax in R
- including plots in homework

Reseach Question

Puzzle

- Does sharing a common gender unite or divide opposing partisans who identify as women?

Hypotheses:

1. **Opposing partisans who identify as women conceive of their gender identity differently.**
2. Mistrust increases among partisans who identify as women when the salience of gender increases.

Source

- Klar, Samara. 2018. "When Common Identities Decrease Trust: An Experimental Study of Partisan Women." American Journal of Political Science. 62(3): 610-622.

The Data

```
load("commonidentity.Rdata")
colnames(df)
```

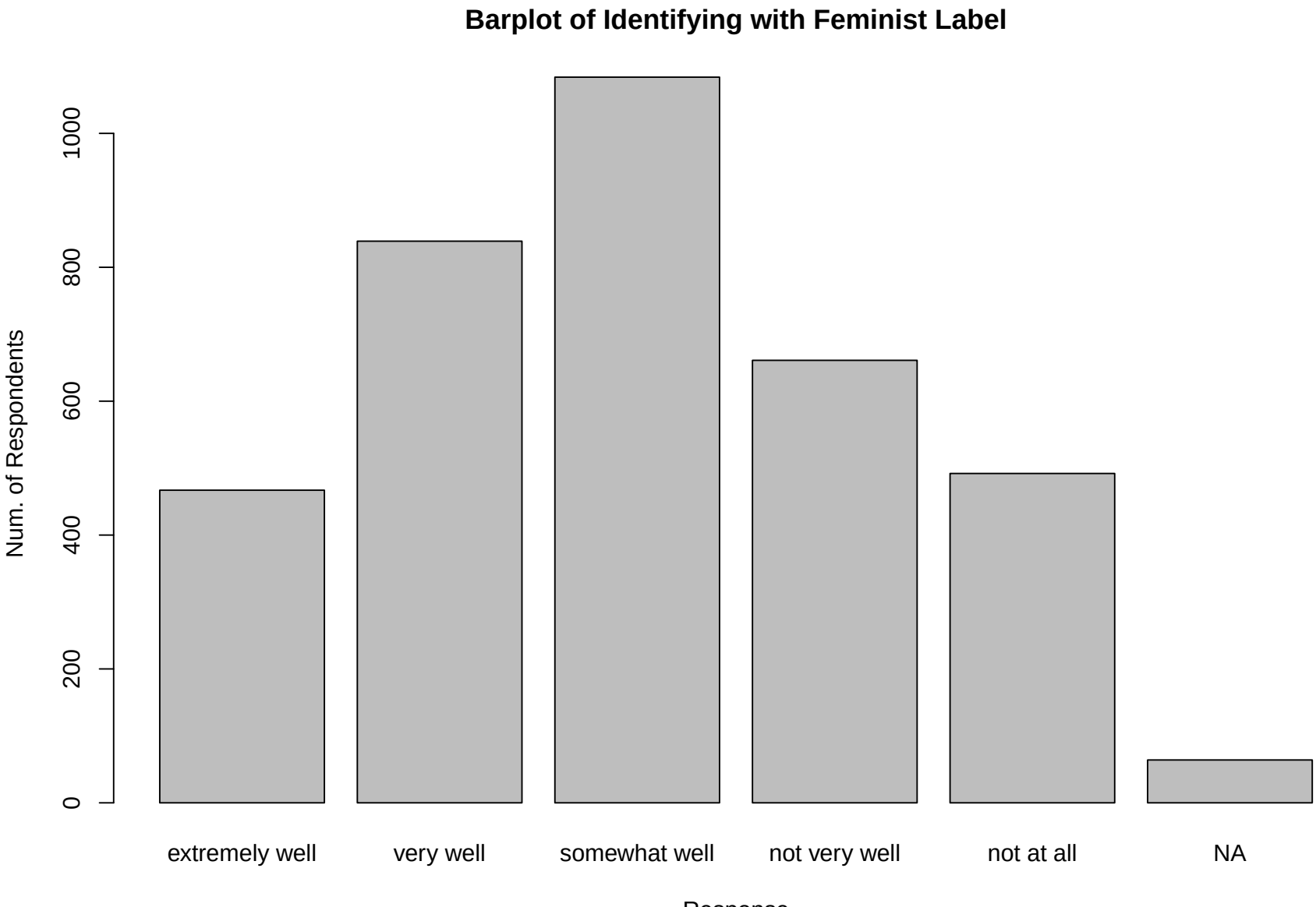
```
##  [1] "feminist_describe_you"        "feminist_democratic_women"
##  [3] "feminist_describe_republican" "ideology"
##  [5] "party_id"                     "party_lean"
##  [7] "age"                          "race"
##  [9] "education"                    "state"
```

- `feminist` variables on 5 point scale: extremely well (1), very well (2), somewhat well (3), not very well (4), not at all (5)
- To test hypothesis #1, we'll focus on `feminist_describe_you` and `party_id`

Hypothesis #1

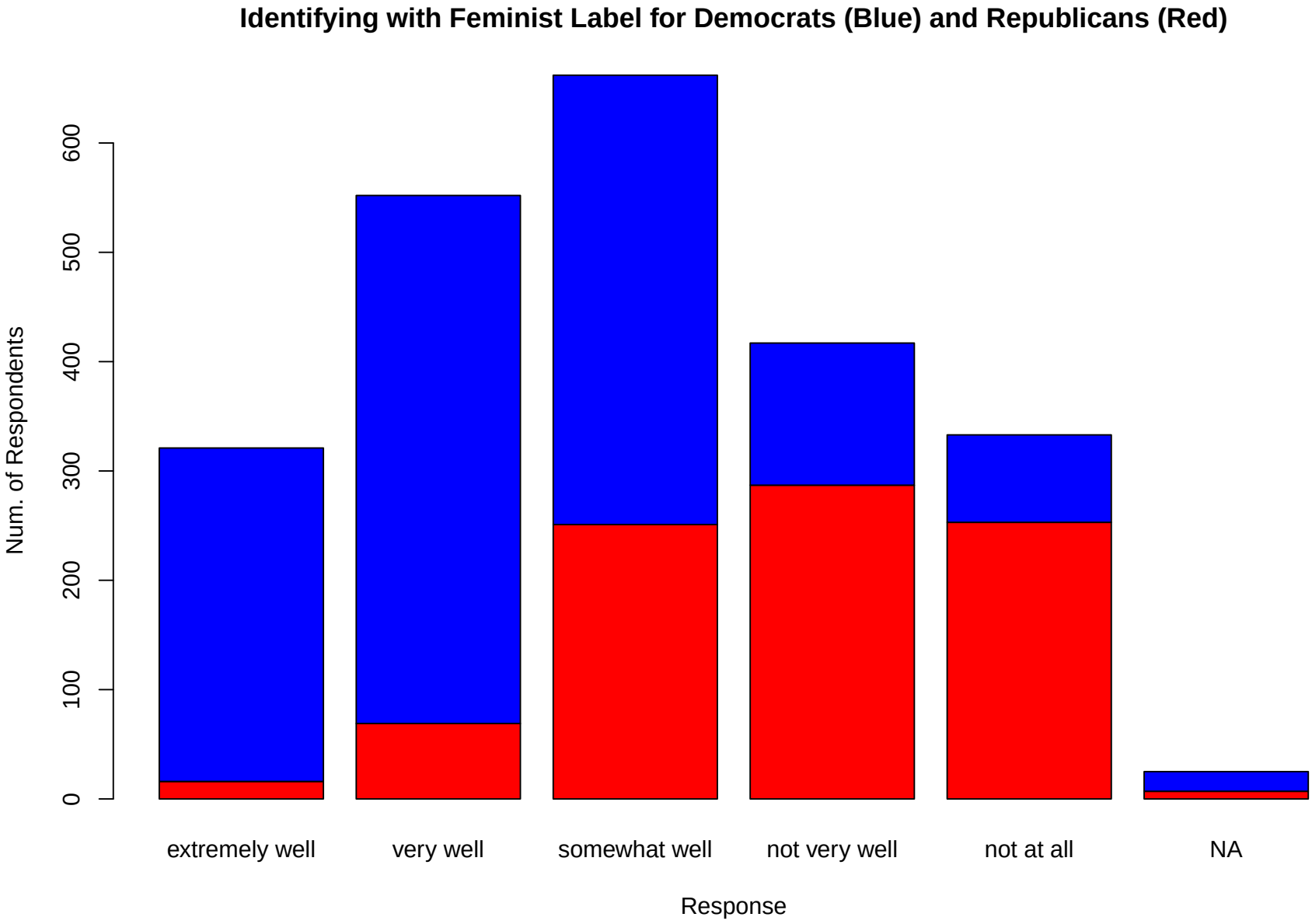
Survey to online panel of 3,607 women aged 18+ April 2017

What does this plot show?



Hypothesis #1

Does visual evidence support or refute hypothesis 1?



Alternative Hypothesis

The confounder of *age* explains both party identification and identifying with the "feminist" label. Perhaps:

- younger people more likely to be democrat
- younger people more likely to identify as feminists

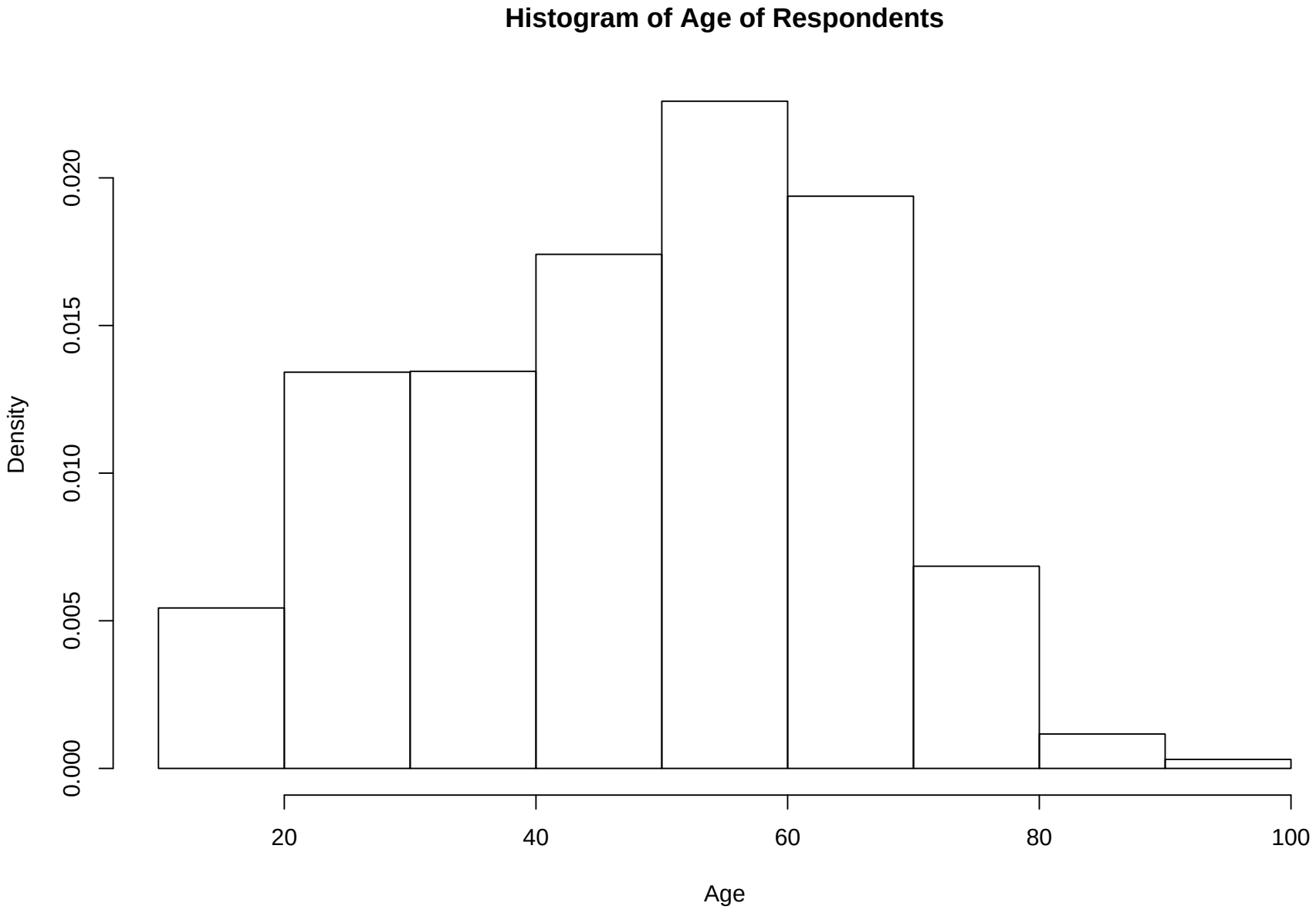
Let's examine with visuals.

Visualizing Data: Histogram

- Histogram shows *density*
- What is density?
- Why does it sum to 100%?

Visualizing Data: Histogram

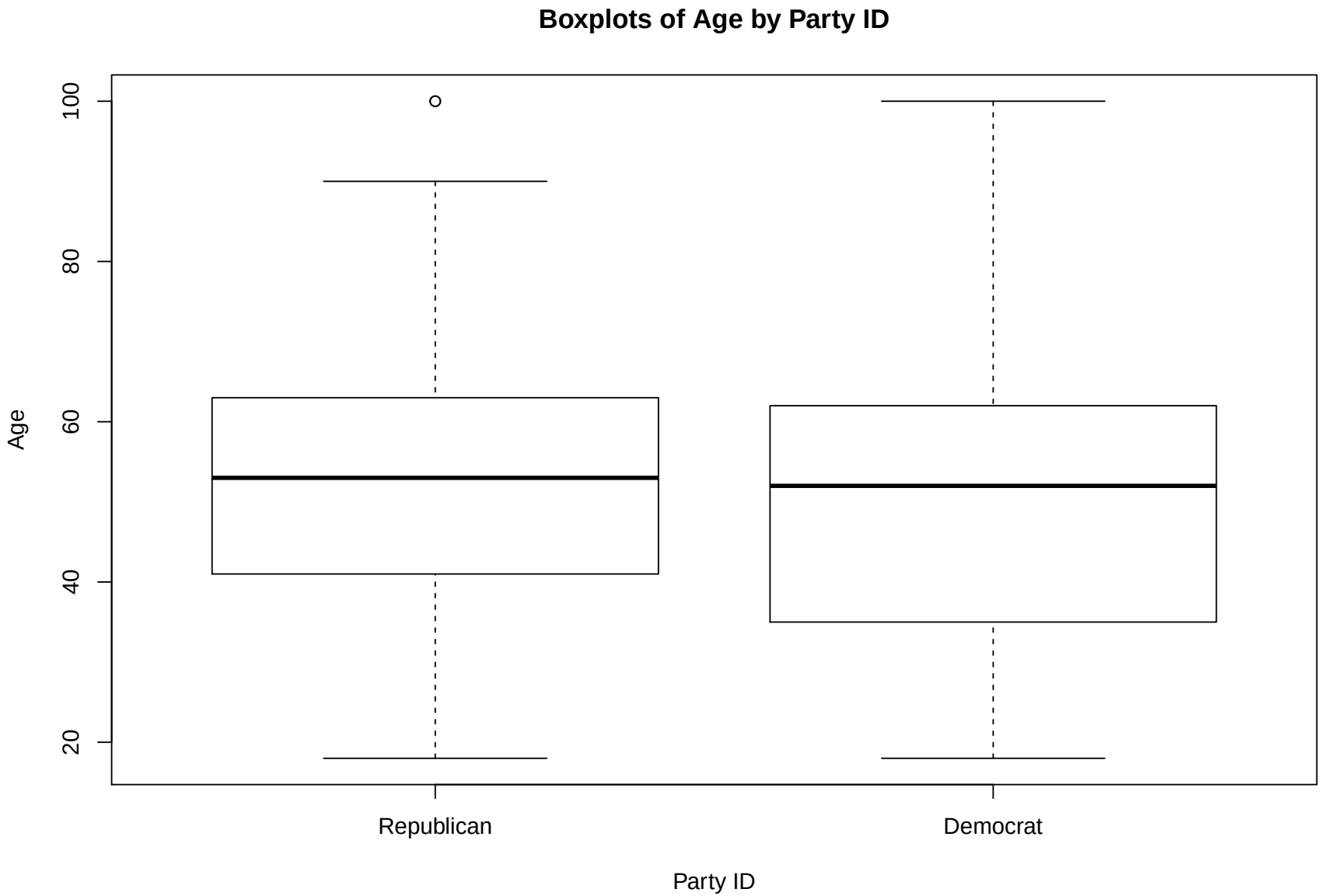
What does this plot show? Interpret first "box."



Visualizing Data: Boxplots

What are the components of a boxplot?

What can we now say about our confounder suspicions?



Visulaizing Data: Correlation

What is correlation?

- Positive: when `x` increases, `y` increases
- Negative: when `x` increases, `y` decreases
- Strong vs. Weak
- Between -1 and 1

Using different data so we have continuous variables.

- Ask, does state spending on education correlate with student success?

Visulaizing Data: Correlation

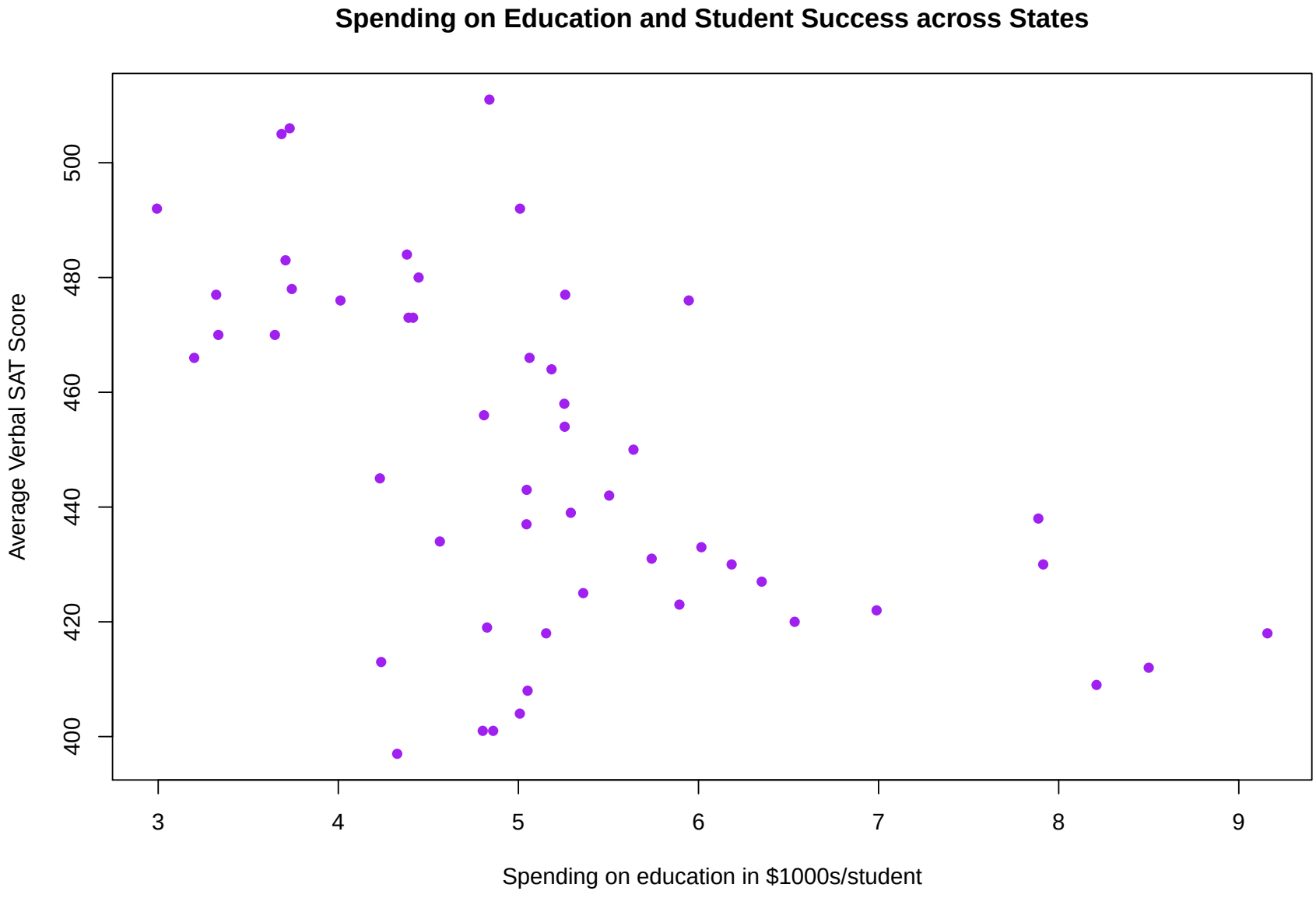
```
#install.packages("car")
library("car")
data("States")
?States ## look at variable descriptions
cor(x = States$dollars, y = States$SATV)
```

```
## [1] -0.5268313
```

```
cor(x = States$dollars, y = States$SATM)
```

```
## [1] -0.4844477
```

Visulaizing Data: Scatterplot



Logarithmic Function

Useful to "smooth out" skewed variable:

- keep in mind for income, population, etc.
- functions in R for this

```
log(3)
```

```
## [1] 1.098612
```

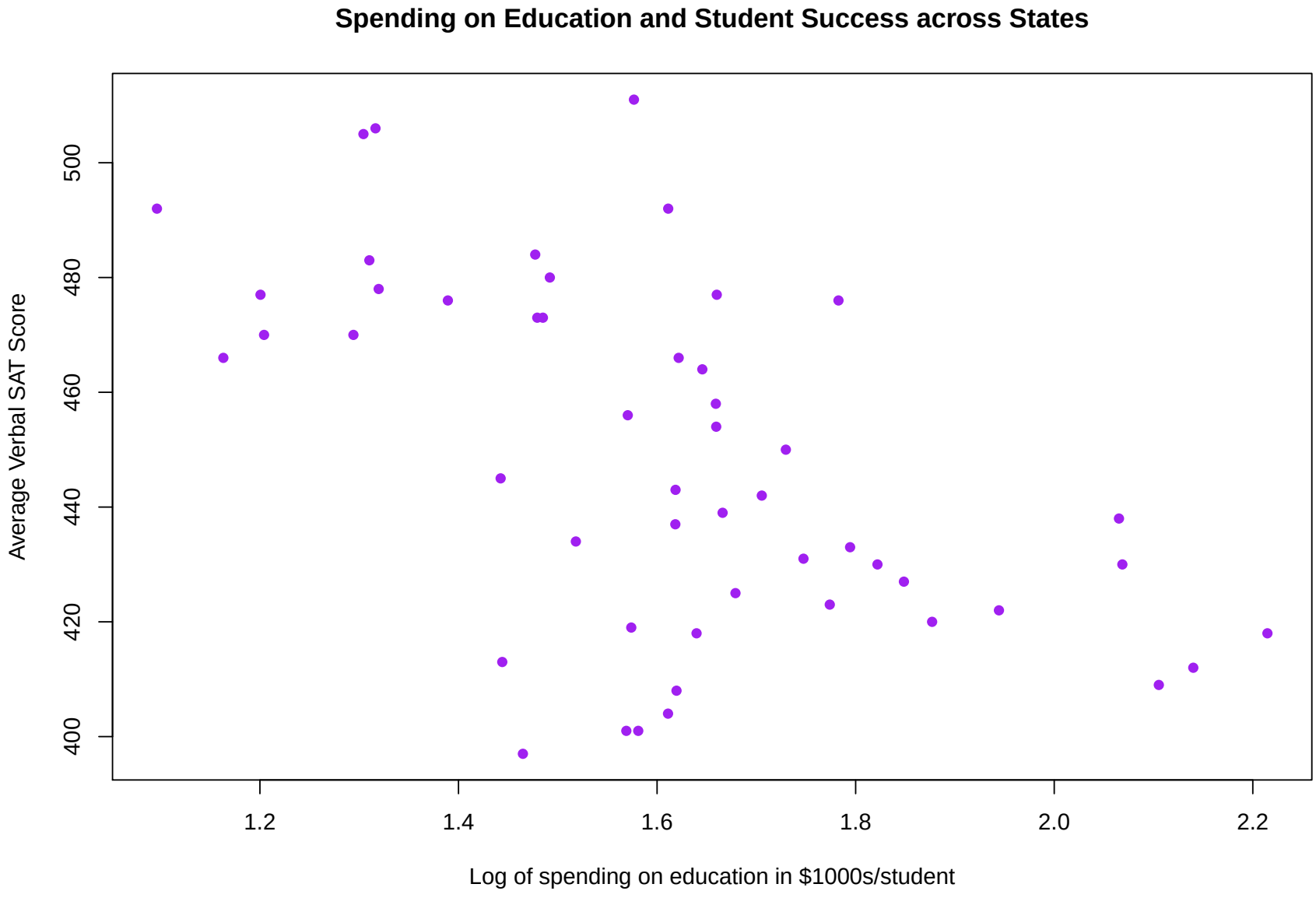
```
log(3, base = exp(1)) ## default base
```

```
## [1] 1.098612
```

```
log(3, base = 10)
```

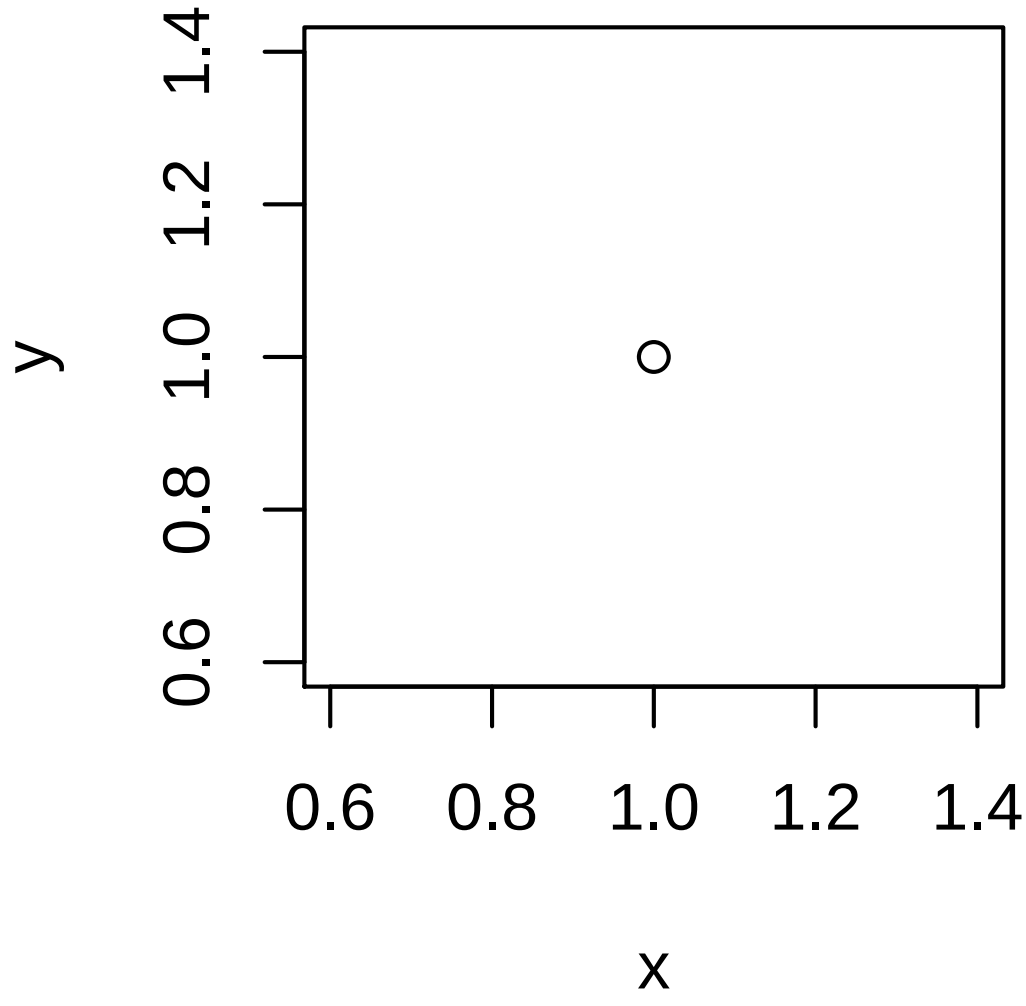
```
## [1] 0.4771213
```

Logged Spending



Plotting in R markdown

To make plots look good in homwork pdf use r chunk options (see .Rmd).



Msc: Tips for Homework Success

- Lots and lots of plotting options... google questions!
- Knit as you go.
- This homework is longer that's why you have 2 weeks...
- Remember, homework 3 we will start to deduct points for formatting if it impacts our ability to grade. Questions?
    1. Write written answers to problems in the text space, not in comments in the R chunks.
    2. Relatedly, make sure lines of code and comments in R chunks do not run off the pdf page.
    3. Make sure that answers from code are displayed in the pdf. If not working, use echo=T and eval=T in the r chunk options, explicitly.