

Class 5

Sean Westwood

Summarize data at every possible value of a variable

We often want to summarize data at every possible value of a variable. For example, we might want to know about GPAs and fraternities on campus. We could create a character variable that holds the name of each fraternity, but converting this character variable to a factor makes computations and visualizations much easier.

```
frats <- c("ΑΦΑ", "ΑΧΑ", "ΒΑΩ", "ΒΓ", "ΓΔΧ", "ΘΔΧ",  
          "ΚΠΚ", "ΛΤΛ", "ΣΝ", "ΣΦΕ", "ΦΔΑ", "ΧΓΕ",  
          "ΧΗ", "ΨΤ", "ΖΨ")
```

1. Let's assume we have a group of 45 Dartmouth students that are affiliated with the fraternities below. Cast this list as a factor and print the levels.

```
frat <- c("ΑΦΑ", "ΑΧΑ", "ΒΑΩ", "ΒΓ", "ΓΔΧ", "ΘΔΧ",  
         "ΚΚΚ", "ΛΤΛ", "ΣΝ", "ΣΦΕ", "ΦΔΑ", "ΧΓΕ",  
         "ΧΗ", "ΨΤ", "ΖΨ", "ΑΦΑ", "ΑΧΑ", "ΒΑΩ", "ΒΓ",  
         "ΓΔΧ", "ΘΔΧ", "ΚΠΚ", "ΛΤΛ", "ΣΝ", "ΣΦΕ",  
         "ΦΔΑ", "ΧΓΕ", "ΧΗ", "ΨΤ", "ΖΨ", "ΑΦΑ", "ΑΧΑ",  
         "ΒΑΩ", "ΒΓ", "ΓΔΧ", "ΘΔΧ", "ΚΠΚ", "ΛΤΛ", "ΣΝ",  
         "ΣΦΕ", "ΦΔΑ", "ΧΓΕ", "ΧΗ", "ΨΤ", "ΖΨ")
```

2. Let's randomly generate some GPAs for each of the 45 students. Compute the mean GPA for all students. This will be different for each of you!

```
frat.gpa <- as.data.frame(frat)  
frat.gpa$GPA <- runif(45, 0, 4)
```

3. In this hypothetical we now want to calculate the average GPA for each fraternity. We could do this in many ways. Some that would just suck because of tons of repetitive typing:

```
mean(frat.gpa$GPA[frat.gpa$frat == "A+A"])
```

```
[1] 2.773724
```

```
mean(frat.gpa$GPA[frat.gpa$frat == "AXA"])
```

```
[1] 1.232255
```

4. Because `frat` is a variable with set levels, we can have R compute the mean GPA for each `frat` in a single command: *tapply()*

`tapply()` expects at least three arguments:

argument	purpose
X	The variable of interest
INDEX	The factor variable you want to group observations by
FUN	The function you want to apply (no parentheses)

Using `tapply()`, compute the mean GPA for each `frat`

Descriptive Statistics for a Single Variable

Continuous variables

Can take any numeric value between a minimum and maximum:

- War casualties: 0 to full population
- Votes: 0 to all eligible voters
- Beers: 0 to all the beers

When describing continuous variables, we may be interested in:

The total (`sum`)

```
sum(frat.gpa$GPA)
```

```
[1] 90.70467
```

The mean (average)

```
mean(frat.gpa$GPA)
```

```
[1] 2.015659
```

The range (the minimum and maximum values)

```
range(frat.gpa$GPA)
```

```
[1] 0.04256854 3.86070732
```

The median

```
median(frat.gpa$GPA)
```

```
[1] 2.058291
```

Percentiles or quantiles: dividing the data in to even groups

```
summary(frat.gpa$GPA)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.04257	0.77562	2.05829	2.01566	3.16471	3.86071

```
IQR(frat.gpa$GPA)
```

```
[1] 2.389084
```

Spread: The spread of a distribution (deviation of each observation from its mean)

On average, how far away are data points from their mean?

$$\text{standard deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

```
sd(frat.gpa$GPA)
```

```
[1] 1.265546
```

Practice

One longstanding debate in the study of international relations concerns the question of whether individual political leaders can make a difference. Some emphasize that leaders with different ideologies and personalities can significantly affect the course of a nation. Others argue that political leaders are severely constrained by historical and institutional forces. Did individuals like Hitler, Mao, Roosevelt, and Churchill make a big difference? The difficulty of empirically testing these arguments stems from the fact that the change of leadership is not random and there are many confounding factors to be adjusted for.

In this exercise, we consider a *natural experiment* in which the success or failure of assassination attempts is assumed to be essentially random.

This exercise is based on: Jones, Benjamin F, and Benjamin A Olken. 2009. “[Hit or Miss? The Effect of Assassinations on Institutions and War.](#)” *American Economic Journal: Macroeconomics* 1(2): 55–87.

```
leaders <- read.csv("leaders.csv")
```

Name	Description
country	The name of the country
year	Year of assassination
leadername	Name of leader who was targeted
age	Age of the targeted leader
politybefore	Average polity score during the 3 year period prior to the attempt
polityafter	Average polity score during the 3 year period after the attempt
civilwarbefore	1 if country is in civil war during the 3 year period prior to the attempt, or 0
civilwarafter	1 if country is in civil war during the 3 year period after the attempt, or 0
interwarbefore	1 if country is in international war during the 3 year period prior to the attempt
interwarafter	1 if country is in international war during the 3 year period after the attempt,
result	Result of the assassination attempt, one of 10 categories described below

The polity variable represents the so-called *polity score* from the Polity Project. The Polity Project systematically documents and quantifies the regime types of all countries in the world from 1800. The polity score is a 21-point scale ranging from -10 (hereditary monarchy) to 10 (consolidated democracy).

The result variable is a 10 category factor variable describing the result of each assassination attempt.

Question 1

1.1. Examine if there is a difference in the age of leaders between successful and failed attempts. We will define a successful attempt as *any* case where the leader died.

Questions 2

2.1. Now check if success rates are related to a civil or international war within the three years prior to the attempt. Create a new binary variable in the dataframe called `warbefore`. Code the variable such that it is equal to 1 if a country was in either a civil or an international war during the 3 years prior to each assassination attempt. To do this compare `warbefore` for successful and failed attempts.

2.2. What does this suggest?

Question 3

3.1. Does a successful assassination correlate with democratization? To address this compare average polity scores before and after successful assassinations.

3.2. Does successful leader assassination correlate with countries who go to war? Look for differences in both civil and international wars.

Question 4

4.1. What is the average age of leaders for each level of result?

4.2. What is the average polity score before an attempt for each level of result?

Evaluation

1. On a scale ranging between 1 (Too Hard) and 10 (Too Easy), how was today's class:
2. What was the easiest thing to understand?
3. What was the most difficult thing to understand?
4. How long did you spend on the assignment outside of class?
5. Please take a moment and tell me about the how the overall class is going for you (good and bad).