

Stat 313 - First Synchronous Meeting!

Allison Theobald

September 23, 2020

Welcome to Our Course

Group Norms:

- Zero tolerance for: racism, sexism, homophobia, transphobia, ageism, ableism
- Be the expert of your own experience – Use “I” statements
- Respect one another
- Intent and impact *both* matter
- Non-judgmental
- Take space, make space – Share the air
- Embrace discomfort
- Make decisions by consensus

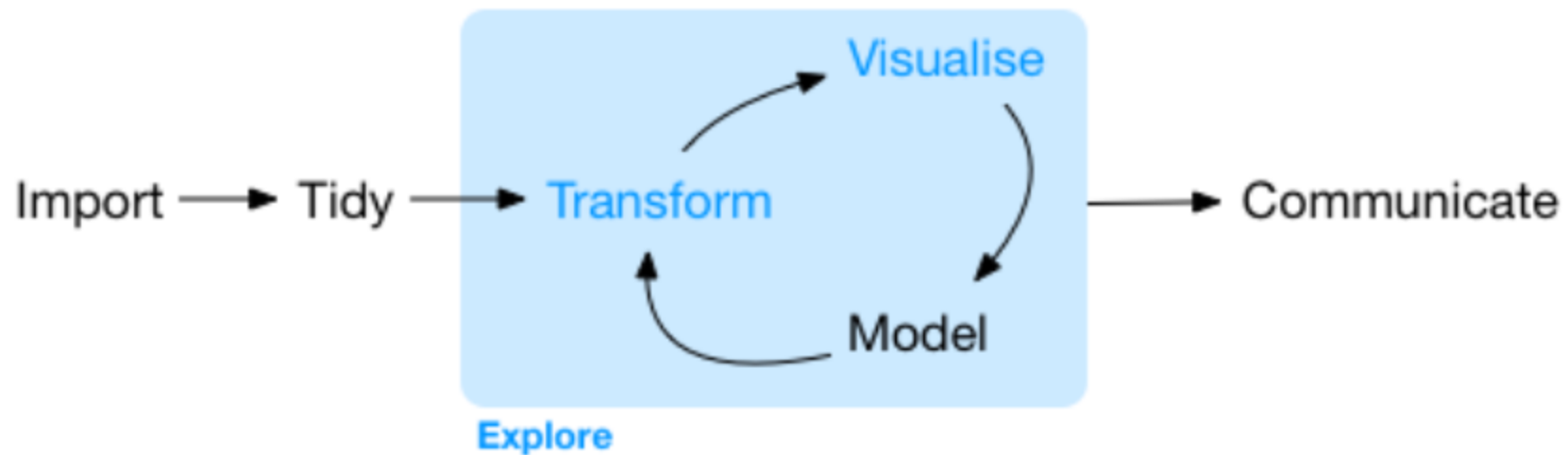
What is Statistics?

Scientists seek to answer questions using rigorous methods and careful observations. These observations – collected from the likes of field notes, surveys, and experiments – form the backbone of a statistical investigation and are called data.

Statistics is the study of how best to collect, analyze, and draw conclusions from data.

Introduction to Modern Statistics

What Statistics Is To Me



Data

Gender stereotypes in 5-7 year old children

subject sex	age trait	target	stereotype	high_achieve_caution
25 male	6 nice	children	0.75	1.00
7 female	6 nice	adults	0.75	0.50
3 male	7 smart	adults	0.25	0.25
70 male	5 smart	adults	0.50	1.00
11 male	6 smart	adults	1.00	0.75
5 male	5 nice	adults	0.25	0.75

In Every Context

Heights of NBA players from the 2008-2009 season

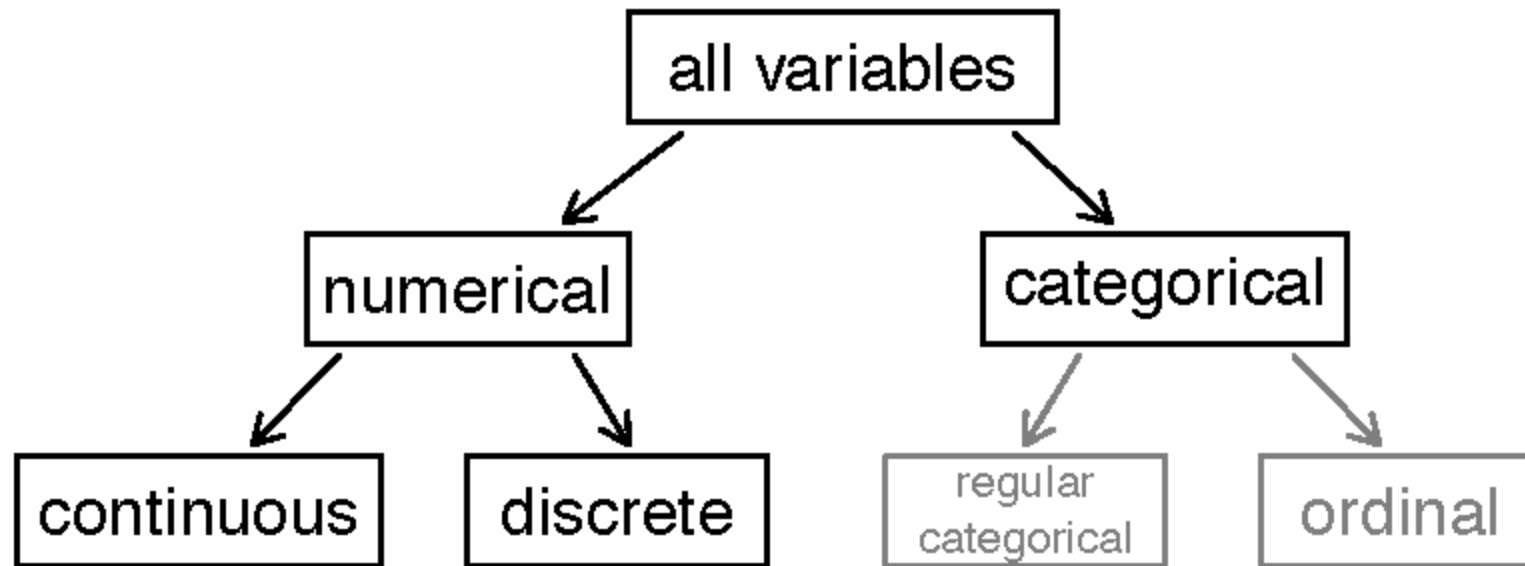
last_name	first_name	h_meters	h_in
West	David	2.06	81
Biedrins	Andris	2.11	83
Garnett	Kevin	2.11	83
Cardinal	Brian	2.03	80
Hunter	Othello	2.03	80
Graham	Stephen	1.98	78

You Could Imagine

Voting history of the United Nations General Assembly

rcid	country	country_code	vote
3798	Thailand	TH	yes
2171	Yugoslavia	YU	yes
4224	Kuwait	KW	yes
4101	Indonesia	ID	abstain
3973	Marshall Islands	MH	yes
4306	Sweden	SE	yes

Types of Variables



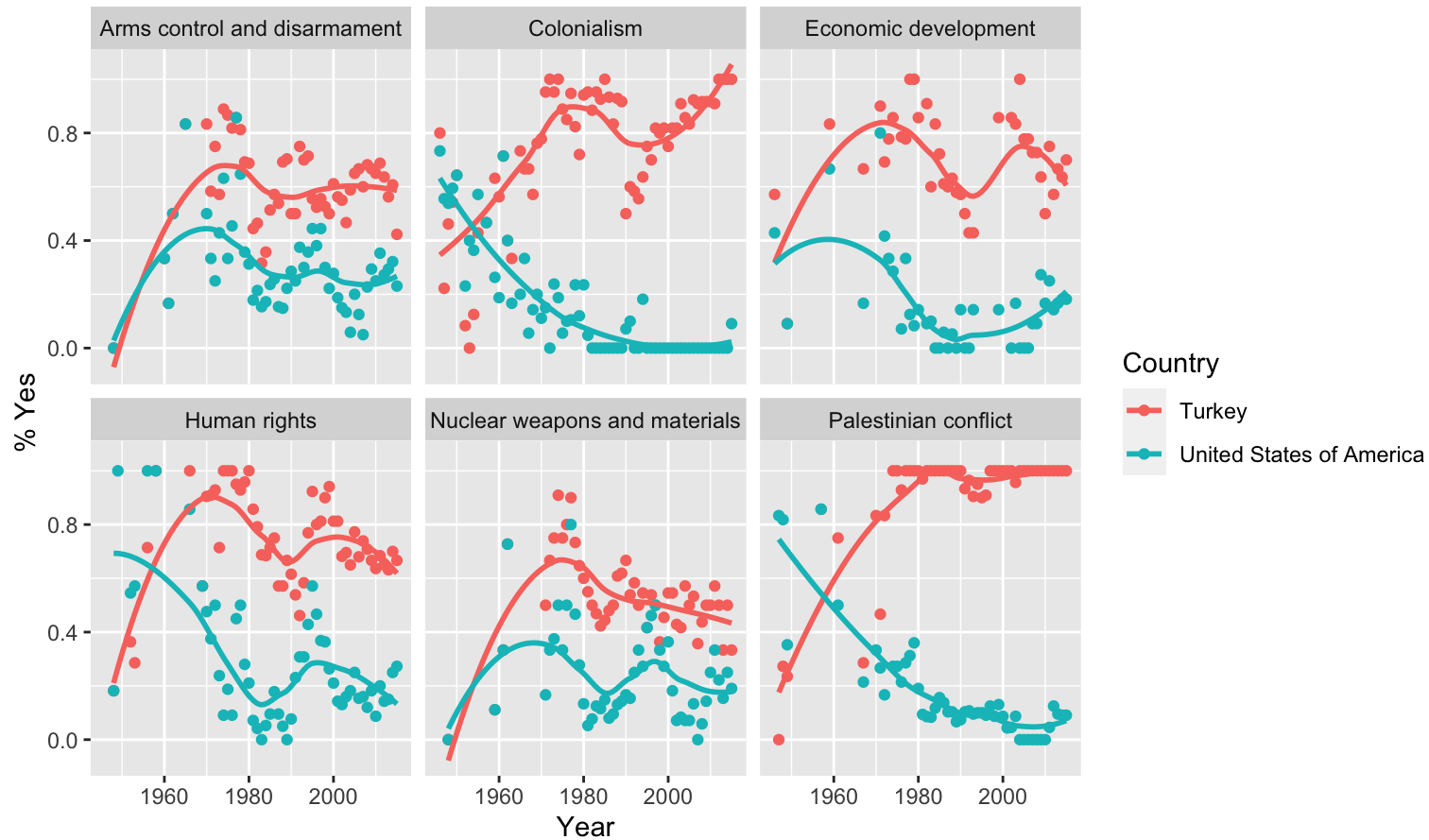
Data Types in R

```
loan50
```

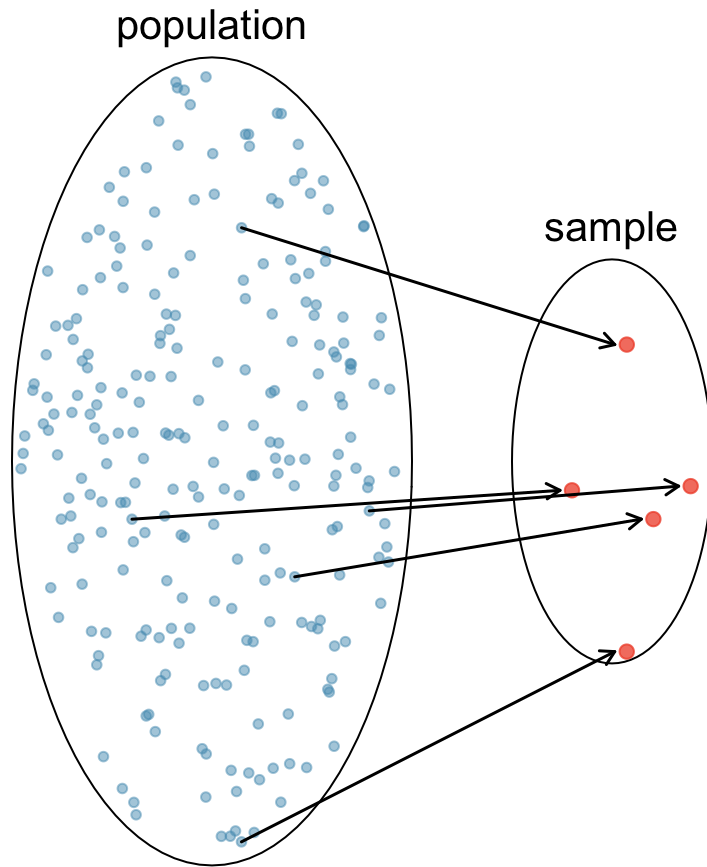
```
## # A tibble: 50 x 6
##   state term ownership income verified_income has_second_income
##   <fct> <dbl> <fct>      <dbl> <fct>          <lgl>
## 1 NJ      60 rent        59000 Not Verified   FALSE
## 2 CA      36 rent        60000 Not Verified   FALSE
## 3 SC      36 mortgage    75000 Verified      FALSE
## 4 CA      36 rent        75000 Not Verified   FALSE
## 5 OH      60 mortgage   254000 Not Verified   FALSE
## 6 IN      36 mortgage    67000 Source Verified FALSE
## 7 NY      36 rent        28800 Source Verified FALSE
## 8 MO      36 mortgage    80000 Not Verified   FALSE
## 9 FL      60 rent        34000 Not Verified   FALSE
## 10 FL     60 mortgage    80000 Source Verified TRUE
## # ... with 40 more rows
```

90 Second Writing Exercise

Percentage of 'Yes' votes in the UN General Assembly, from 1946 to 2015



Sampling Strategies



- Simple random sample
- Stratified random sample
- Cluster sample
- Multistage sample

Types of Studies

Experiment

- randomization
- replication
- controlling
- blocking

Observational Study

- collect data in a way that does not directly interfere with how the data arise

Relationships Between Variables

explanatory variable \rightarrow *might affect* \rightarrow response variable

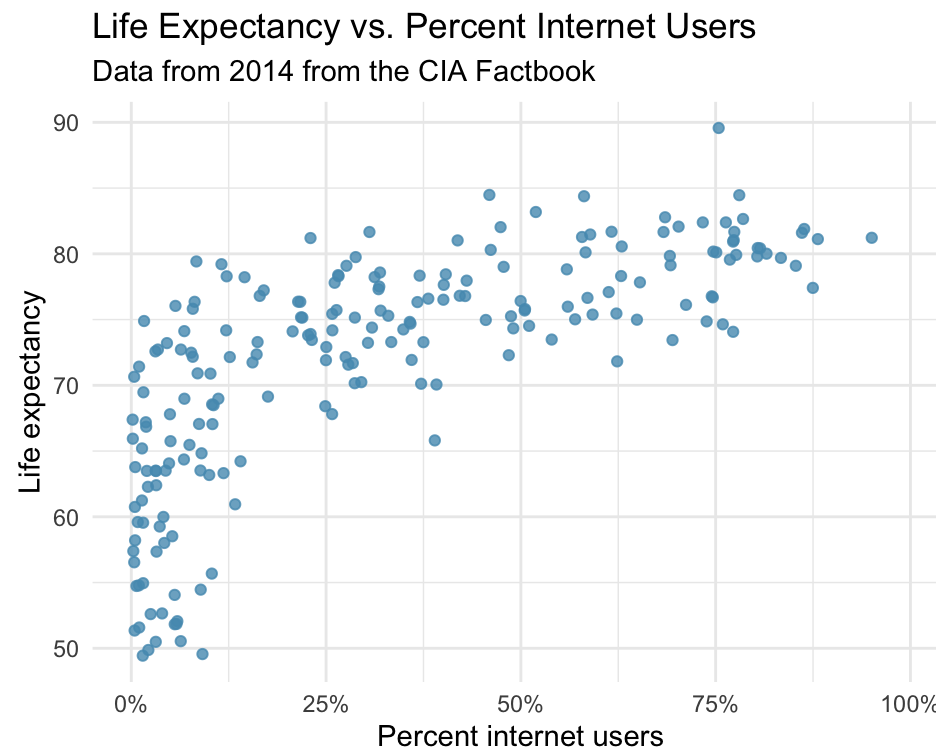
- If two variables are not associated, then they are said to be **independent**.
- If two variables are associated, then they are said to be **dependent**.

association \neq causation

What do you need to say that the explanatory variable causes a change in the response variable?

Group Exercise

The following scatterplot was created as part of a study evaluating the relationship between estimated life expectancy at birth (as of 2014) and percentage of internet users (as of 2009) in 208 countries for which such data were available.



Questions for Me

- Content questions?
- Course structure questions?
- Programming in `R` questions?

Drop In **R** Help Session(s)

- Wednesday after class (1:00 - 3:00)
- Thursday office hours (4:00 - 5:00 pm)
- Post to **R** Questions channel in Canvas