What is Econometrics?

EC 320: Introduction to Econometrics

Boyoon Chang
Winter 2022

Prologue

Who am I?

Boyoon (Bo) Chang

- Doctoral student in economics
- Former research associate in economics team at a law firm
- Focus in applied microeconomics, empirical industrial organization

Where can you find me?

- Office: 420 PLC
- Office hours: MW 15:00-16:00, or by appointment
- Email: bchang@uoregon.edu **(use EC 320 in the subject line)**

Today's Topic

Syllabus

- Course material
- What, when, where, who

Econometrics

- Motivation
- Examples

R

- What is R?
- Why are we using R?
- Getting started with R

Motivation

Why study econometrics?

1. Develop **skills that employers value**.
2. Cultivate **healthy skepticism**.
3. Learn about the world using **data**.

Motivation

Why study econometrics?

Provide answers to important questions

- Do minimum wage policies **reduce poverty**?
- Does the death penalty **deter violent crime**?
- Are recessions **good for your health**?
- How will global warming **affect the economy**?
- What **explains the gender pay gap**?

Econometrics

Most econometric inquiry concerns one of two distinct goals:

1. **Prediction:** Accurately predict or forecast an outcome given a set of predictors. Given what we know about x, what values do we expect y to take?

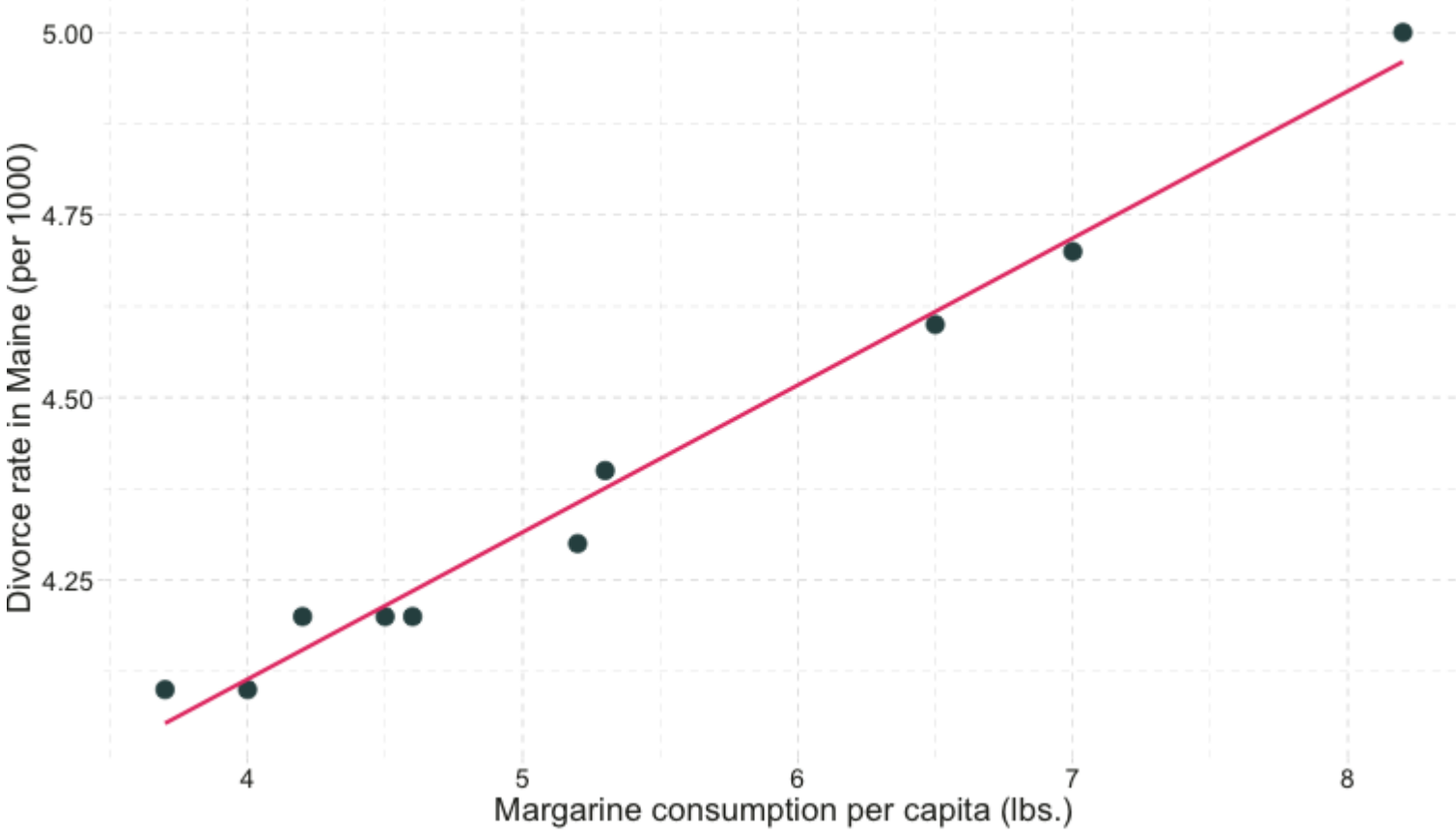
2. **Causal identification:** Estimate the effect of an intervention on an outcome. How does y change when we change x?

The main focus of EC 320 and EC 421 is causal identification.

- But...both rely on a common set of statistical techniques.

- For those interested, Professor Tim Duy teaches forecasting (EC 422) this Winter.

Not all relationships are causal



Correlation vs. Causation

Common refrain: *"Correlation doesn't necessarily imply causation!"*

- **Q:** Why might correlation fail to describe a causal relationship?
- **A:** Omitted-variables bias, selection bias, simultaneity, reverse causality.

Correlation can imply causation.

- Requires strong assumptions.
- **Real life often violates these assumptions!**
- **Solutions:** Conduct an experiment or find a natural experiment.

Example: *Blue Paradox*

Recent study by UO economist Grant McDermott and coauthors.

Question: Do commercial fishers preempt fishing bans by increasing their fishing effort before the bans go into effect?

Motivation

- Recent conservation efforts seek to preserve aquatic habitat and increase fish stocks.
- Policy lever: Restrict fishing activity in marine protected areas.
- Concern: Preemptive behavior could *decrease* fish stocks.

Data

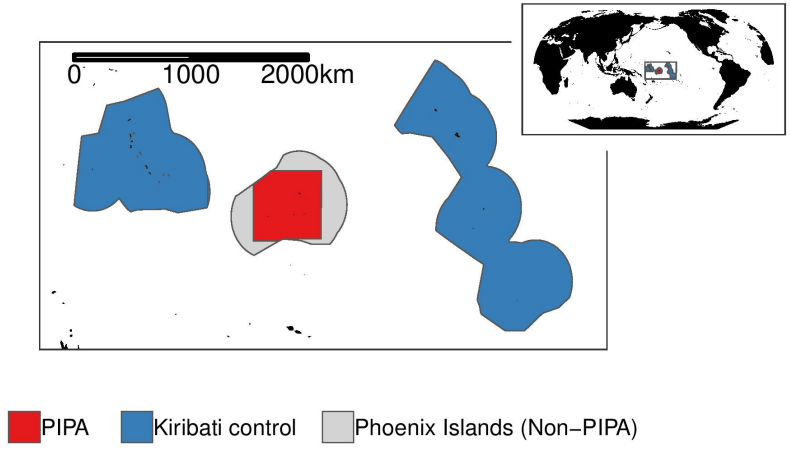
- Vessel-level data on fishing effort/intensity.

Example: *Blue Paradox*

Natural Experiment

Phoenix Islands Protected Area (PIPA)

- First mentioned on 1 September 2014; implemented 1 January 2015.
- *Treatment group:* PIPA.
- *Control group:* Outlying Kiribati islands.



Example: *Blue Paradox*

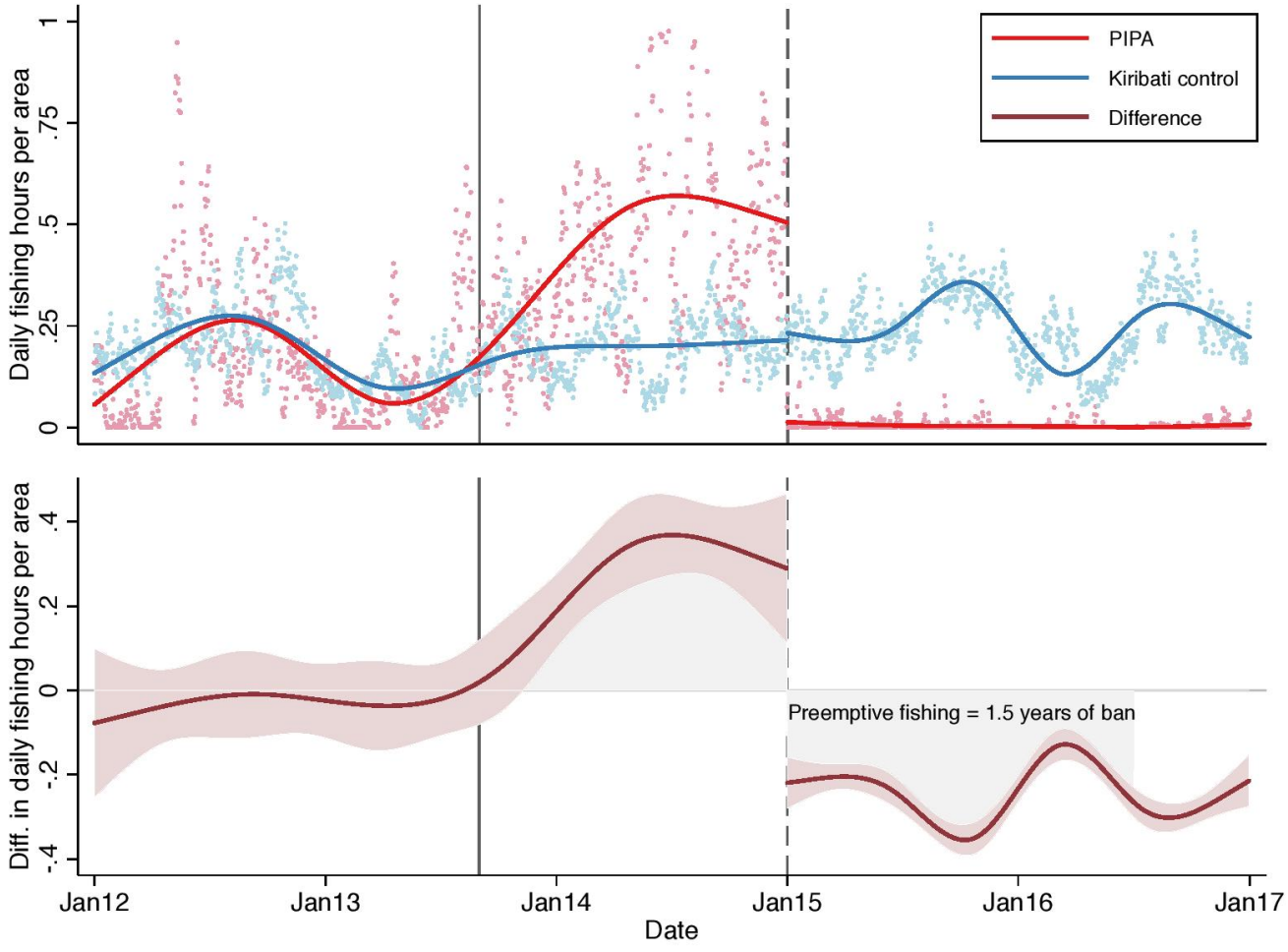
Natural Experiment

Measure the causal effect of the fishing ban by comparing fishing effort in treatment and control regions, before-and-after PIPA.

- A *difference-in-differences* comparison.
- **Assumption:** Parallel trends. If we believe this assumption, then the observed change supports a causal interpretation. If not, then the change could reflect other factors and thus fail to isolate the causal effect of the ban.

Example: *Blue Paradox*

Results



Example: *Blue Paradox*

Discussion

Results provide causal evidence that commercial fishers engage in preemptive behavior in response to conservation policy changes.

Results are *consistent* with economic theory, but *cannot prove* that the theory is correct.

- **Science cannot prove anything.**
- Science can **falsify or reject** existing hypotheses or **corroborate** existing evidence.

Also...the causal statement rests on a critical assumption.

- Cannot prove that the assumption is true, but can falsify it.
- Failure to falsify $\neq$ assumption is true.

R

What is R?

According to the R project website,

> R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

What does that mean?

- R is **free** and **open source**.
- R executes a variety of statistical techniques and produces beautiful graphs.
- R has a vibrant, thriving online community (see stack overflow).

Why are we using R?

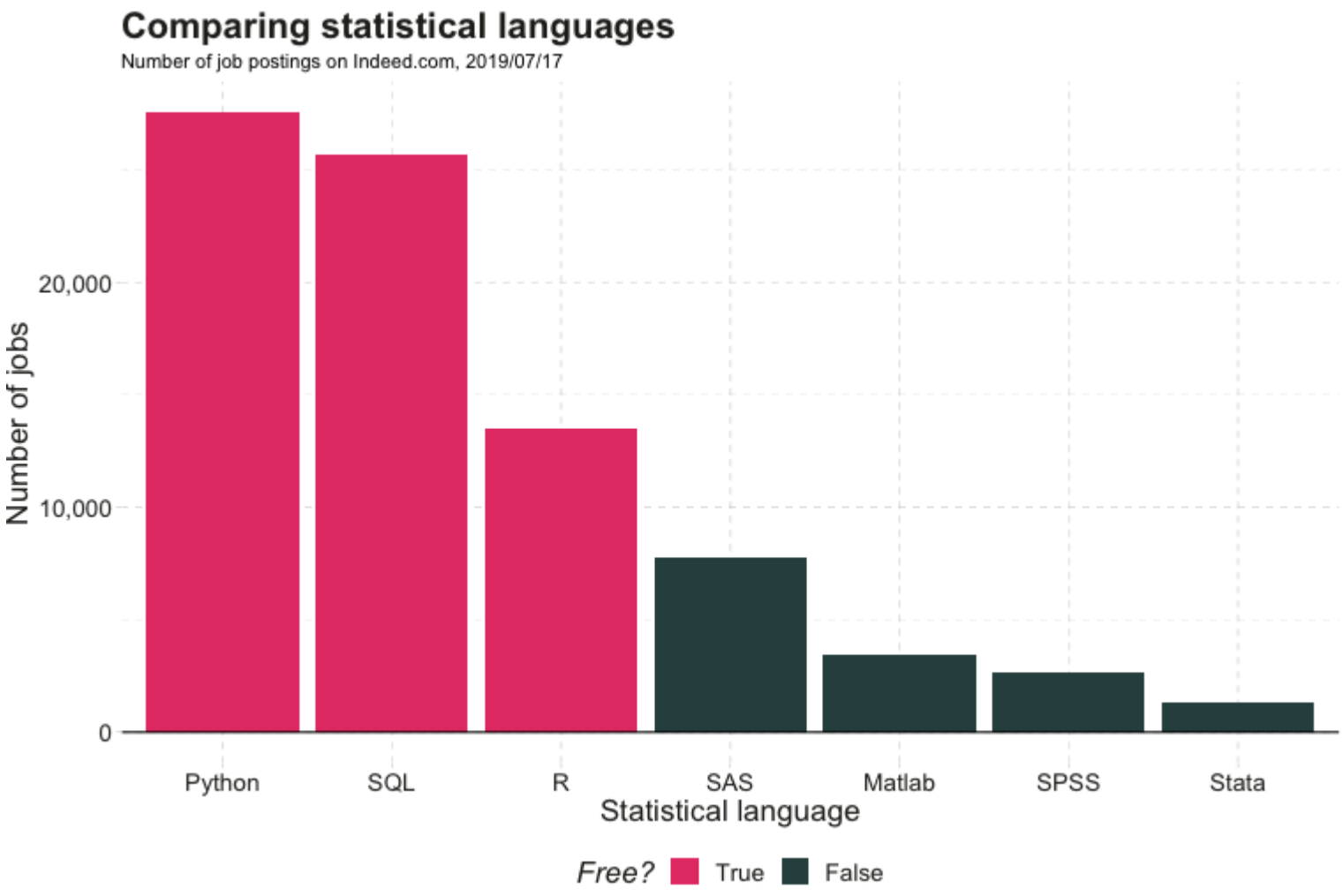
1. R is **free**.

2. **R is popular** among economists, political scientists, psychologists, sociologists, geographers, anthropologists, biologists, data scientists, and statisticians.

3. **Employers prefer R** over most competing software environments.

4. R can **adapt to nearly any task**: 'metrics, spatial data analysis, machine learning, web scraping, data cleaning, website building, teaching.

R


Comparing statistical languages
Number of job postings on Indeed.com, 2019/07/17
Number of jobs
20,000
10,000
0
Python
SQL
R
SAS
Matlab
SPSS
Stata
Statistical language
Free?
True
False

R + [Examples]

R + Regression

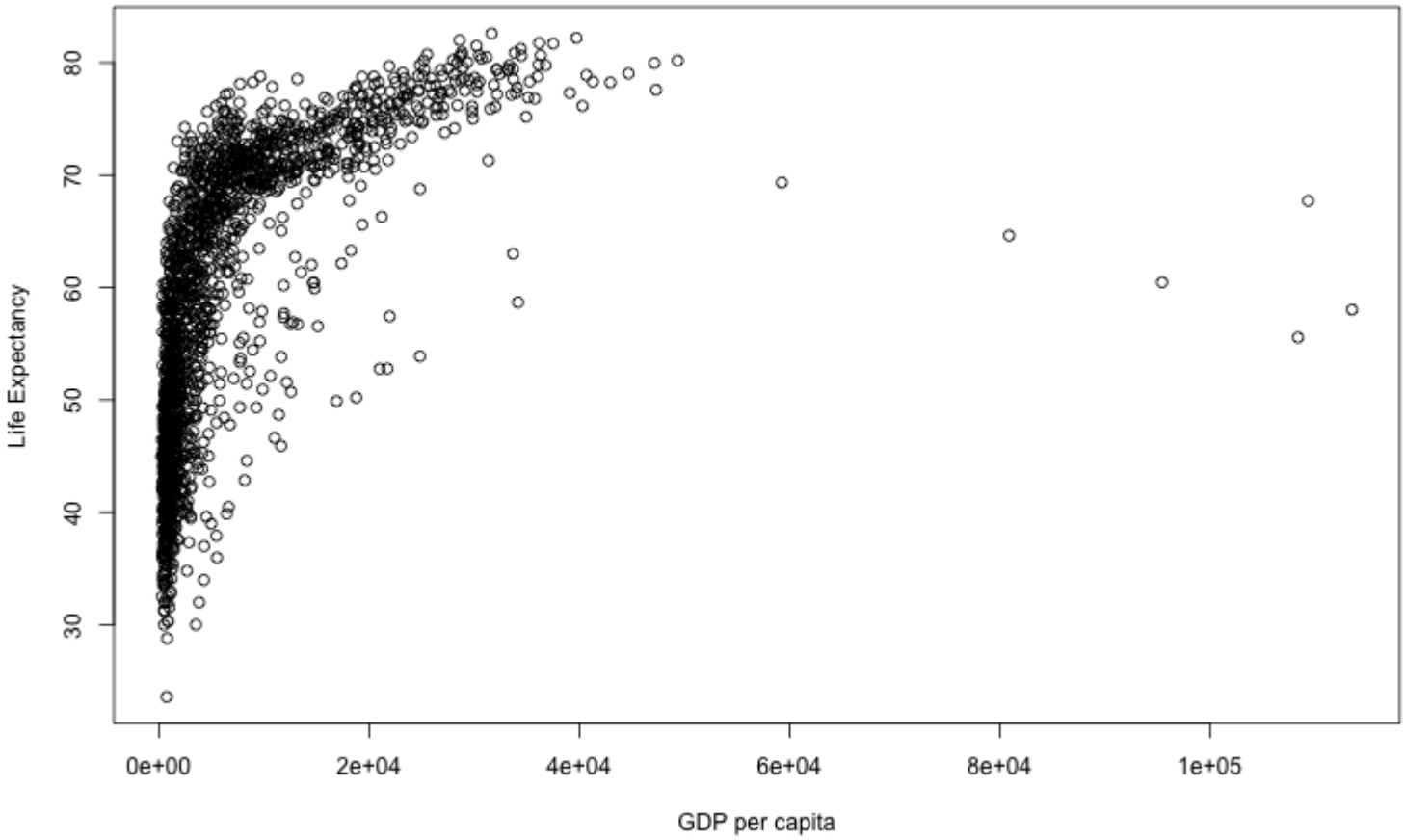
```r
# A simple regression
fit ← lm(mpg ~ 1 + wt, data = mtcars)
# Show the coefficients
coef(summary(fit))
```

```
#>              Estimate Std. Error   t value     Pr(>|t|)
#> (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
#> wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

```r
# A nice, clear table
library(broom)
tidy(fit)
```

```
#> # A tibble: 2 × 5
#>   term        estimate std.error statistic  p.value
#>   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
#> 1 (Intercept)    37.3      1.88      19.9  8.24e-19
#> 2 wt             -5.34     0.559     -9.56 1.29e-10
```

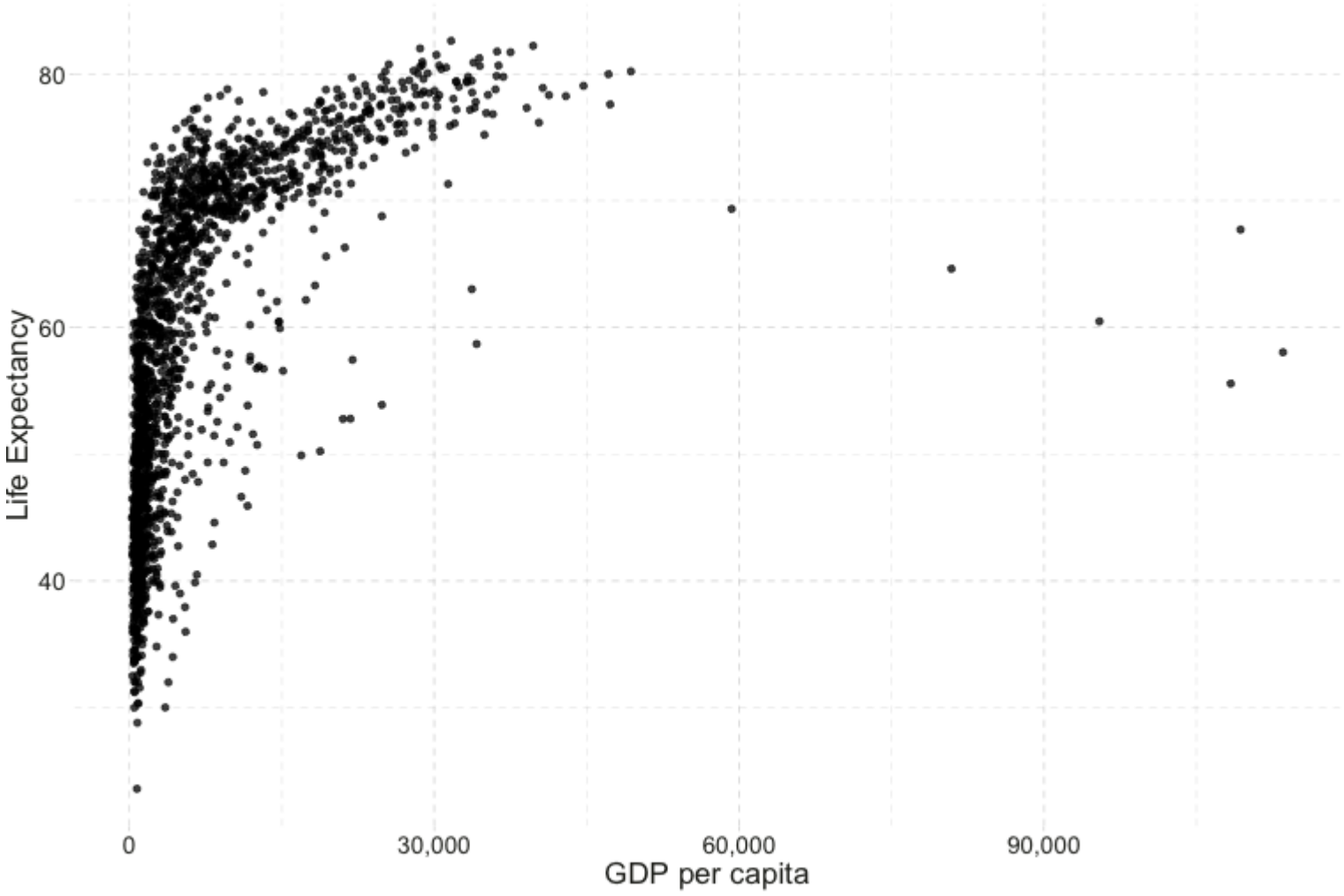
R + Plotting (w/ `plot`)



R + Plotting (w/ plot)

```r
# Load packages with dataset
library(gapminder)
# Create dataset
plot(
  x = gapminder$gdpPercap, y = gapminder$lifeExp,
  xlab = "GDP per capita", ylab = "Life Expectancy"
)
```

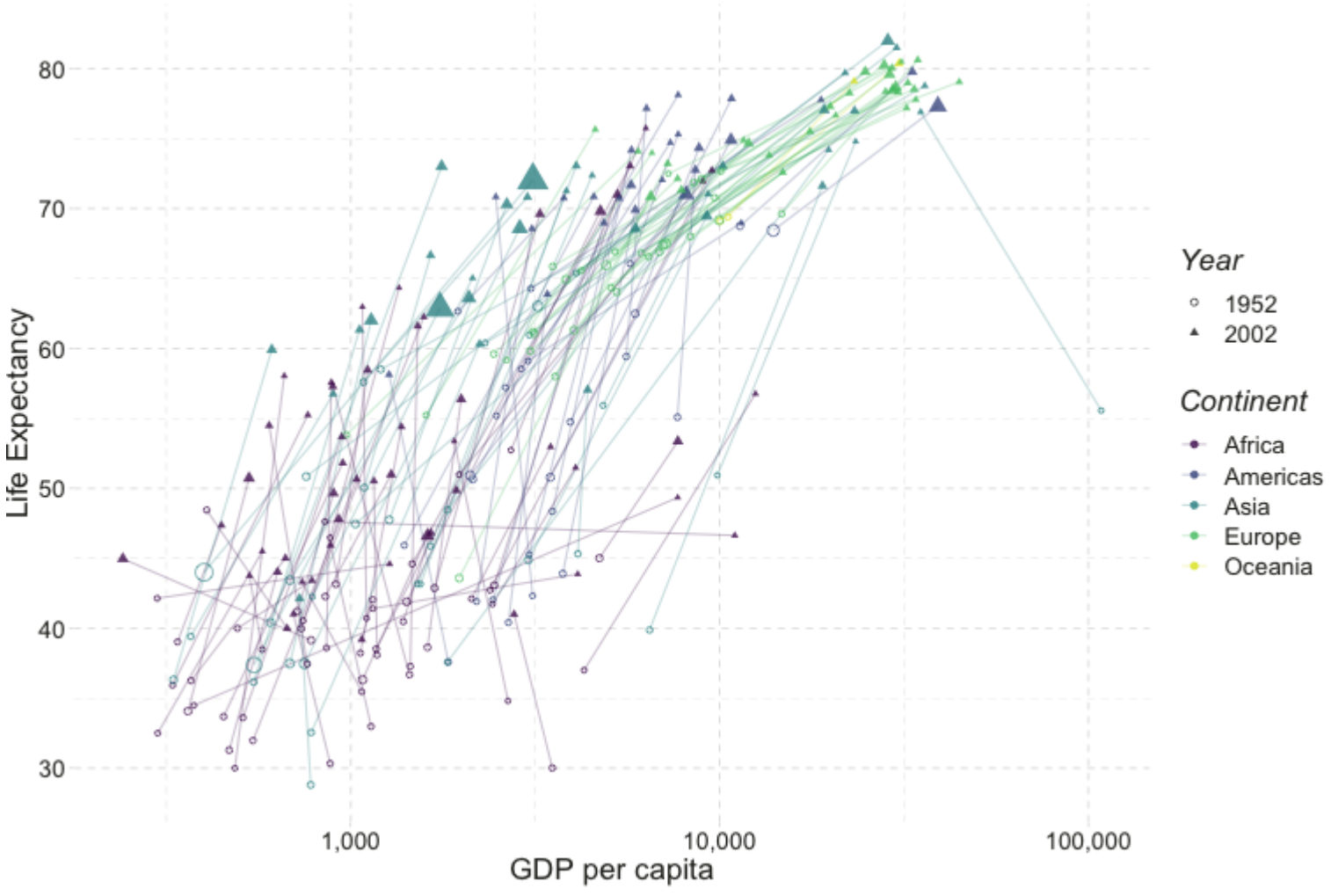
R + Plotting (w/ ggplot2)



R + Plotting (w/ ggplot2)

```r
# Load packages
library(gapminder); library(dplyr)
# Create dataset
ggplot(data = gapminder, aes(x = gdpPercap, y = lifeExp)) +
geom_point(alpha = 0.75) +
scale_x_continuous("GDP per capita", label = scales::comma) +
ylab("Life Expectancy") +
theme_pander(base_size = 17, base_family = "Arial", fc = met_slate)
```

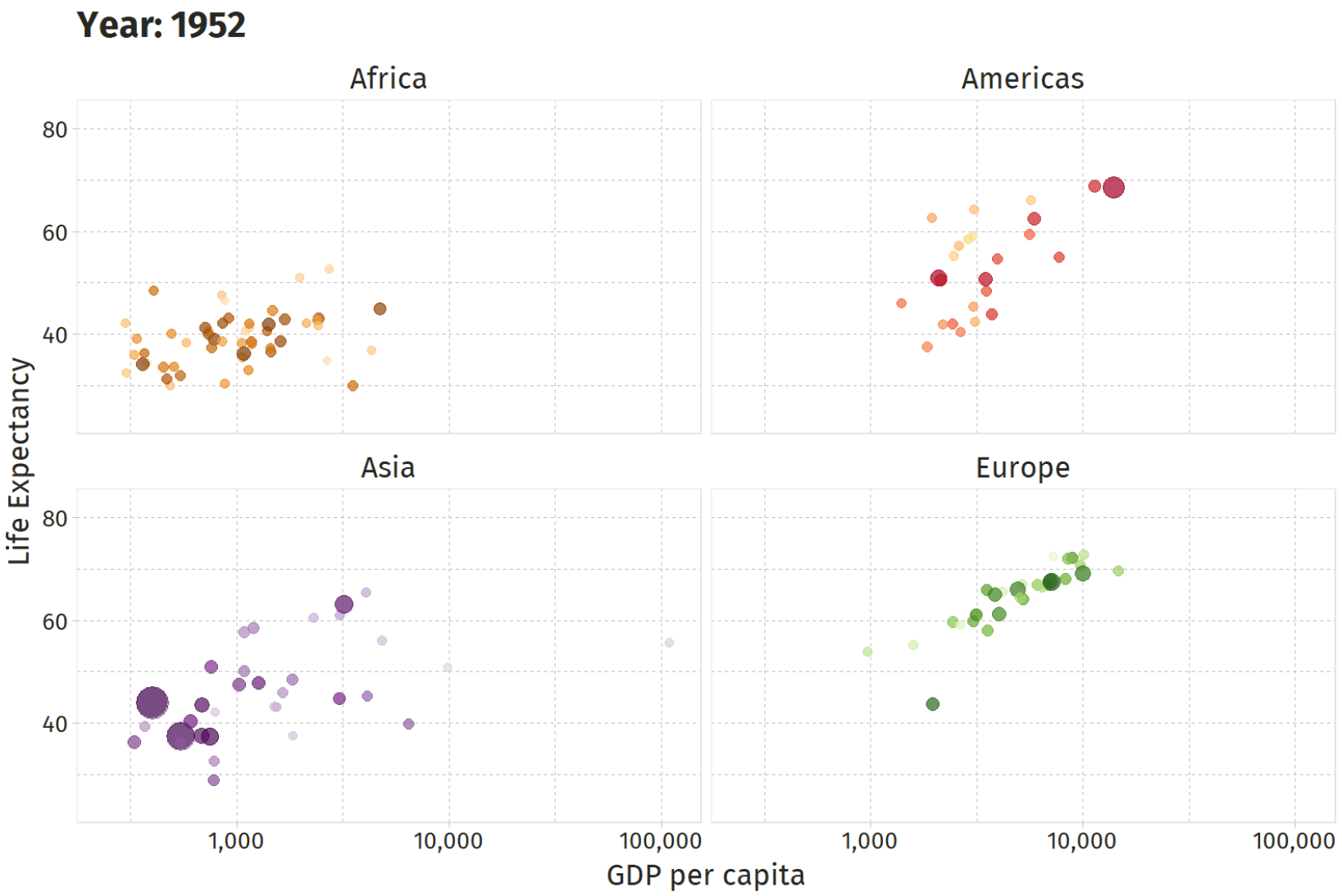
R + More plotting (w/ ggplot2)



R + More plotting (w/ ggplot2)

```r
# Load packages
library(gapminder); library(dplyr)
# Create dataset
ggplot(
  data = filter(gapminder, year %in% c(1952, 2002)),
  aes(x = gdpPercap, y = lifeExp, color = continent, group = country)
) +
geom_path(alpha = 0.25) +
geom_point(aes(shape = as.character(year), size = pop), alpha = 0.75) +
scale_x_log10("GDP per capita", label = scales::comma) +
ylab("Life Expectancy") +
scale_shape_manual("Year", values = c(1, 17)) +
scale_color_viridis("Continent", discrete = T, end = 0.95) +
guides(size = F) +
theme_pander(base_size = 17, base_family = "Arial", fc = met_slate)
```

R + Animated plots (w/ gganimate)



R + Animated plots (w/ gganimate)

```r
# The package for animating ggplot2
library(gganimate)
# As before
ggplot(
  data = gapminder %>% filter(continent != "Oceania"),
  aes(gdpPercap, lifeExp, size = pop, color = country)
) +
geom_point(alpha = 0.7, show.legend = FALSE) +
scale_colour_manual(values = country_colors) +
scale_size(range = c(2, 12)) +
scale_x_log10("GDP per capita", label = scales::comma) +
facet_wrap(~continent) +
theme_pander(base_size = 17, base_family = "Arial", fc = met_slate) +
theme(panel.border = element_rect(color = "grey90", fill = NA)) +
# Add gganimate code
labs(title = "Year: {frame_time}") +
ylab("Life Expectancy") +
transition_time(year) +
ease_aes("linear")
```

R + Animated maps (w/ gganimate)

Getting Started with R

Starting R

Installation

- Install R.
- Install `RStudio`.
- **Note:** All academic workstations at the UO have R, but having a copy of R on your computer will prove useful for the econometrics sequence and 400-level elective courses.

Resources

- Google and StackOverflow
- Time
- Your classmates
- Your GE
- Me

Starting R

R basics

1. Everything is an **object**. `foo`
2. Every object has a **name** and **value**. `foo ← 2`
3. You use **functions** on these objects. `mean(foo)`
4. Functions come in **libraries** (**packages**). `library(dplyr)`
5. R will try to **help** you. `?dplyr`
6. R has its **quirks**. `NA; error; warning`