POLI 30 D: Political Inquiry

TA Sessions

Lab 10 | What can I do with R?

Before we start

Announcements:

- GitHub page: **https://github.com/umbertomig/POLI30Dpublic**
- Piazza forum: Check with instructors for an alternative link.

Before we start

Announcements: Final Exam

- Three questions.
  - Everyone needs to do Question 1.
  - Then you choose between Q2 and Q3.
- Each question: five points (letters a to e)
  - Q1: Theoretical (interpretation)
  - Q2 or Q3: Coding (R) + Theory (interpretation)
- We cannot help you during the exam. Keep calm and put your brain to work.
- You are fantastic. You got this!

Before we start

Recap: In the Lab sessions so far, you learned:

- How to install R and R Studio on your computer.
- How to do basic and advanced operations with vectors and data frames.
- How to install packages and work with R Markdown.
- How to create plots, run data analysis, and run regressions.

Great job!

- Do you have any questions about these contents?

Plan for Lab 10

- Cool stuff you can do if you keep learning R:
  - Join two datasets
  - Plotting a Map
  - Doing an Interactive Plot
  - A bit of Text-as-data
  - A bit of Machine Learning
  - Shiny WebApps
  - Scrollytells

Getting started

Getting started

- I prepared a `Lab10.R` script for you to follow the class.
- Use that script!
- Step 0: Install all the needed packages. Then load them.

Join two datasets

Join two datasets

- Join two datasets is something that R does fast and reliably. Consider you have these two datasets in here:

Dataset 1:

```
##      country courts barb2 prsexp2
## 1  Argentina      0 -0.72       1
## 2  Australia      1 -6.91       5
## 3    Austria      1 -4.91       5
## 4 Bangladesh      0  0.78       1
## 5    Belgium      1 -4.62       5
```

Dataset 2:

```
##      country prscorr2 gdpw2
## 1  Australia        4 10.30
## 2    Austria        4 10.10
## 3 Bangladesh        0  8.38
## 4    Belgium        4 10.25
## 5    Bolivia        0  8.58
```

- Note that they have they variable `country` in common, but different variables. Why could that be?

Join two datasets

To join it, we use the `full_join` command:

```
full_join(dat1, dat2)
## Joining, by = "country"
##      country courts barb2 prsexp2 prscorr2 gdpw2
## 1  Argentina      0 -0.72       1       NA    NA
## 2  Australia      1 -6.91       5        4 10.30
## 3    Austria      1 -4.91       5        4 10.10
## 4 Bangladesh      0  0.78       1        0  8.38
## 5    Belgium      1 -4.62       5        4 10.25
## 6    Bolivia     NA    NA      NA        0  8.58
```

- Note that whenever we have no idea of the value, R fills it with NAs.
- And there are several other ways to join datasets. Keep learning!

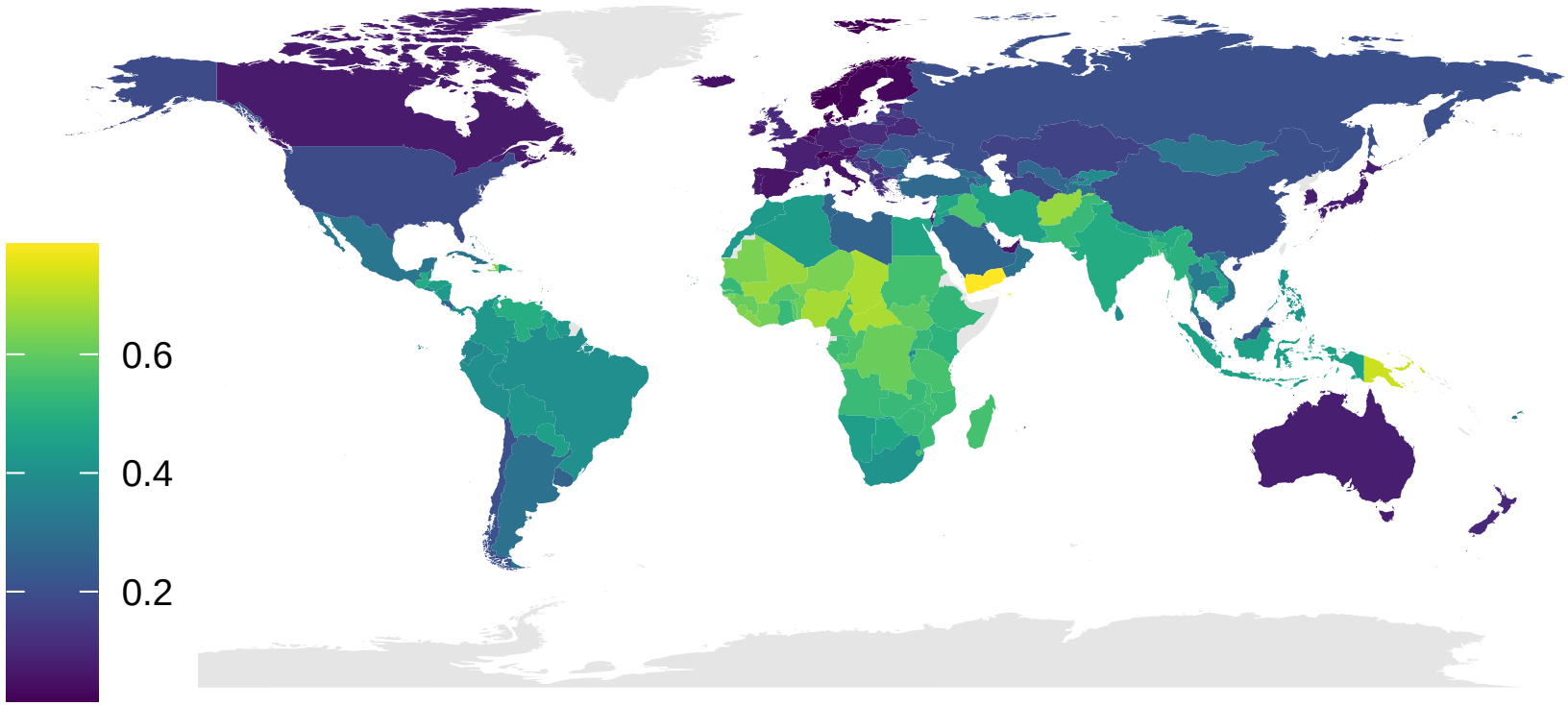
Maps

Maps

- You can plot beautiful (and informative) maps in R.
- There are many ways to do that. We are going to use one here that is the most common.
- We are going to draw two maps:
  1. Gender Inequality Index (2021)
  2. Corruption Perception Index (2018)
- You can do other maps as well. Keep learning!

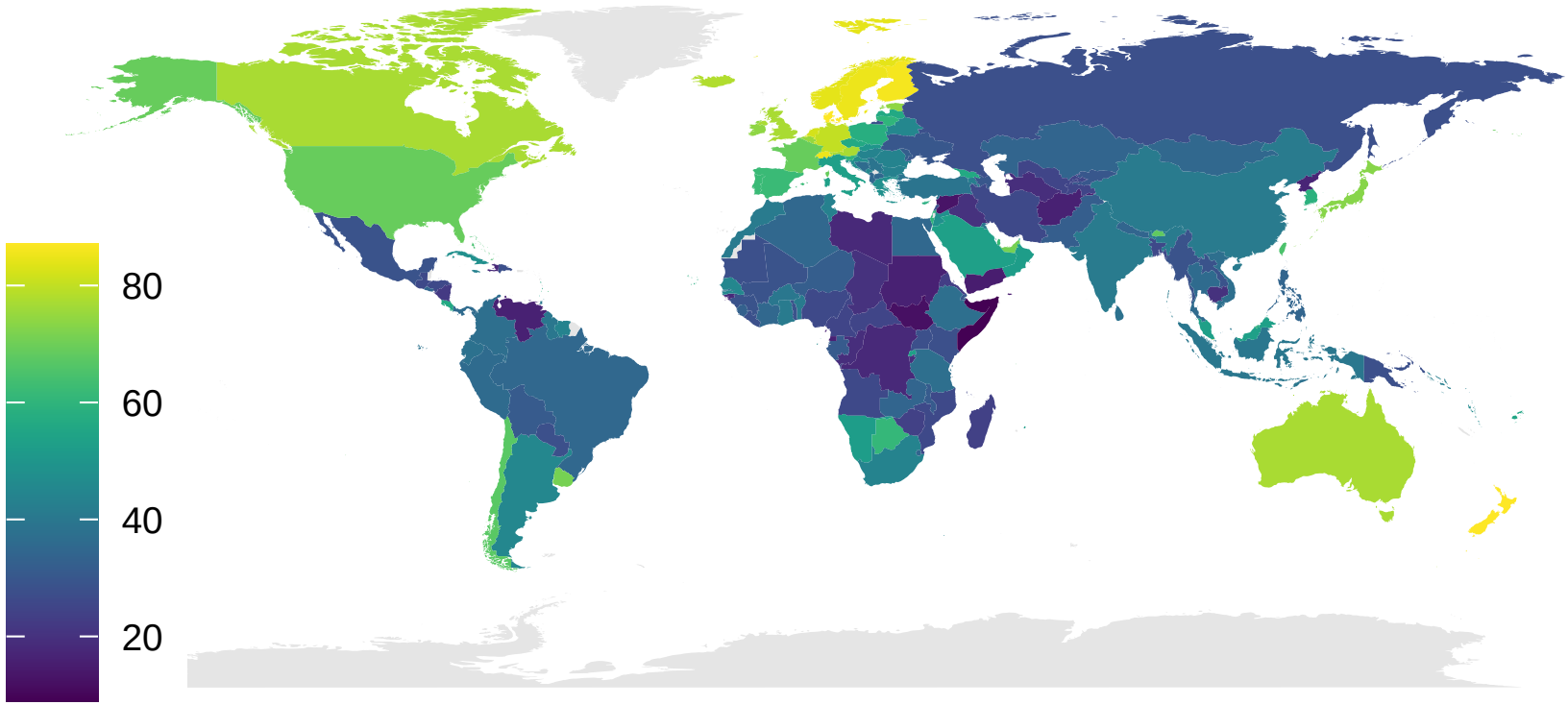
Maps

Gender Inequality Index
(higher values represent more inequality)



Maps

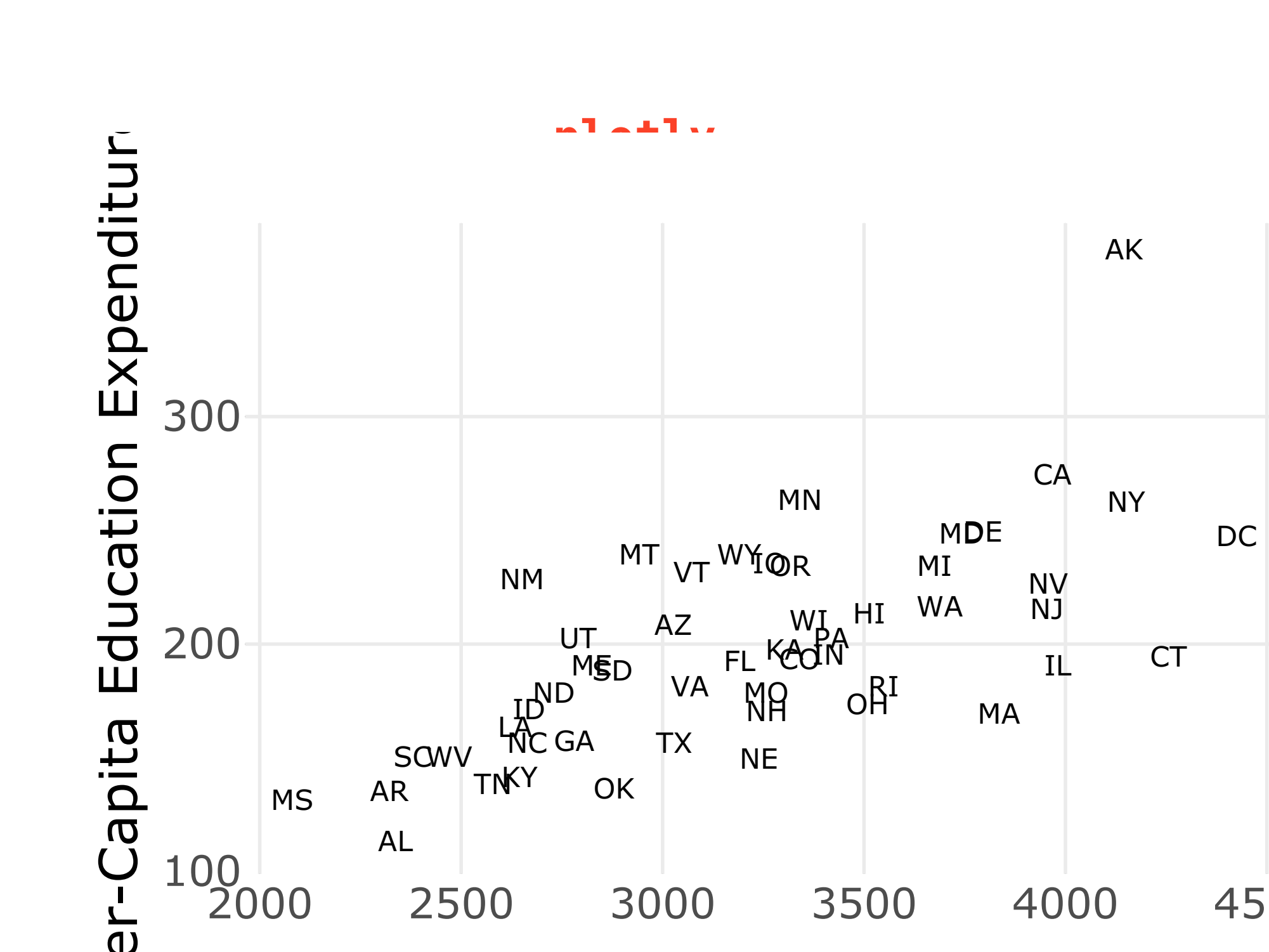
Corruption Perception Index
(higher values represent lower corruption)

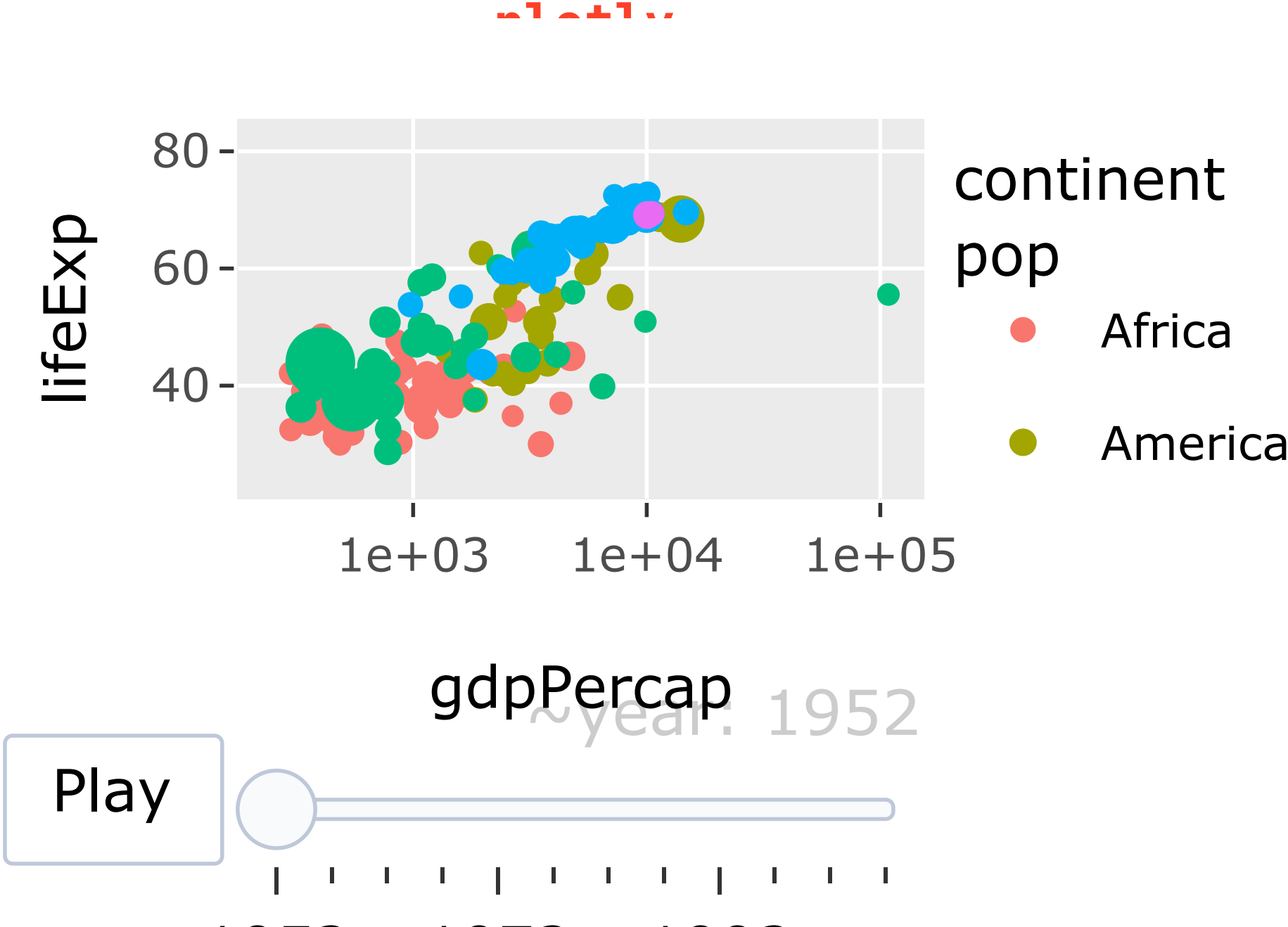


plotly

plotly

- Interactive graphs are great, especially when taking data to people that are not specialists.
- `plotly` helps us to plot the graphs and play with them.
- Since we cannot do it in a PDF, just run the code on your computer.
- You can build your interactive plot. Keep learning!


plotly
Per-Capita Education Expenditur
300
200
100
2000
2500
3000
3500
4000
45
AK
CA
NY
MN
MD
DE
DC
MT
WY
IO
OR
VT
NM
MI
NV
WA
NJ
HI
AZ
WI
PA
UT
KA
IN
CO
ME
SD
FL
IL
CT
ND
VA
MO
RI
ID
NH
OH
MA
LA
NC
GA
TX
NE
SC
WV
TN
KY
OK
AR
MS
AL


plotly
80
60
40
lifeExp
1e+03
1e+04
1e+05
gdpPercap
continent
pop
Africa
America
~year: 1952
Play

Text-as-data

Text-as-data

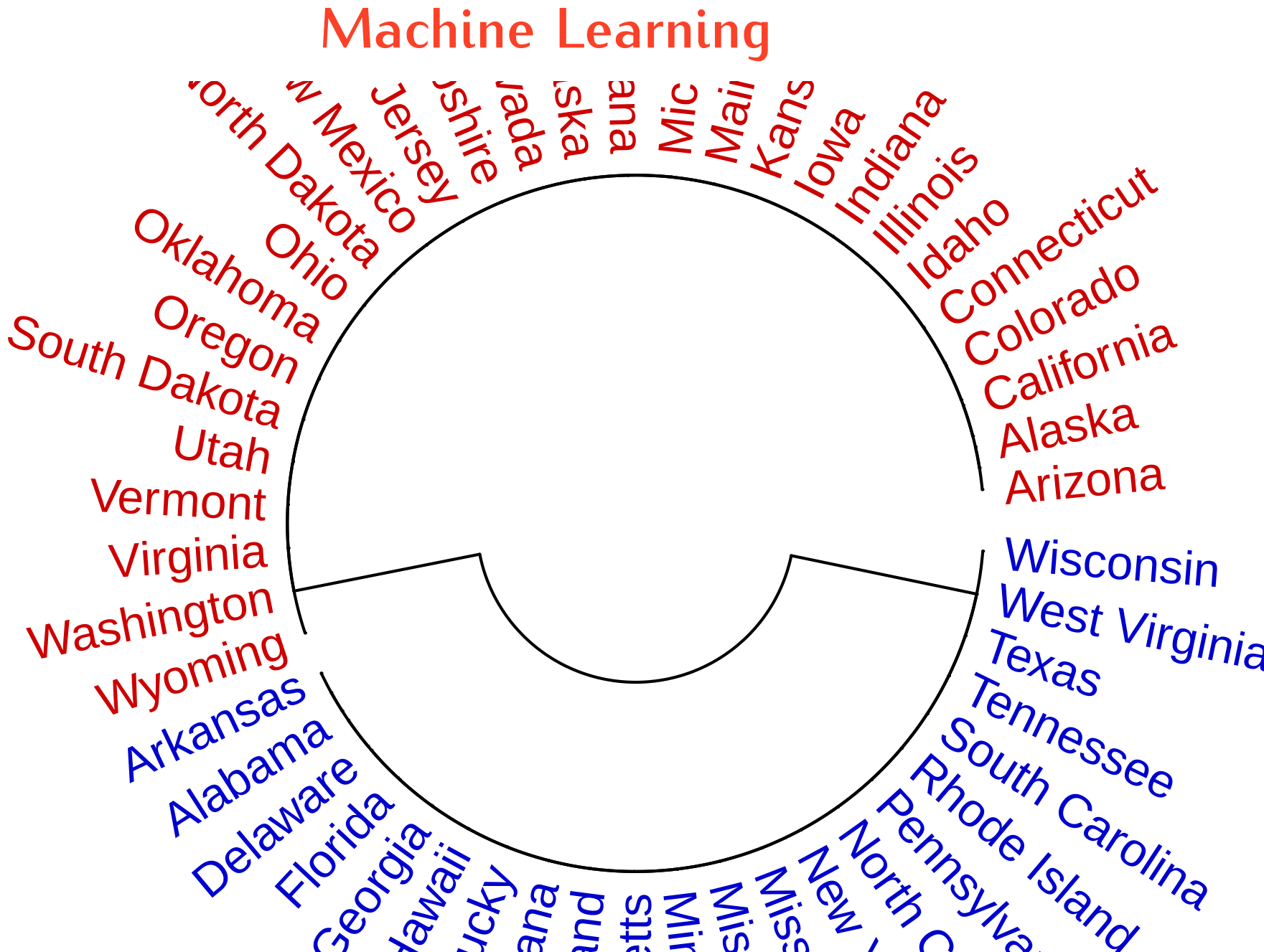
- Lots of our work in PoliSci is about analyzing a massive text corpus.
  - Tweets
  - Books
  - Reports
  - You name it...
- R is at the forefront when it comes to text-as-data.
- You can also analyze text-as-data using R. Keep learning!

oil
said
prices
opec
mln
crude
the
bpd
saudi
dlrs
last
market
reuter
will
official
barrel
price
production
meeting
barrels
pct
one
new
output
futures
industry
kuwait
government
minister
world
billion
petroleum
sheikh
sources
also
month
quota
report
traders
accord
international
contract
analysts
exports
arab
group
growth
gulf
recent
plans
ability
agency
imports
today
april
budget
nazer
nymex
markets
present
economy
emirates
arabia
january
december
energy
exchange
members
sell
says
economic
emergency
help
riyals
they
hold
may
all
but
quoted

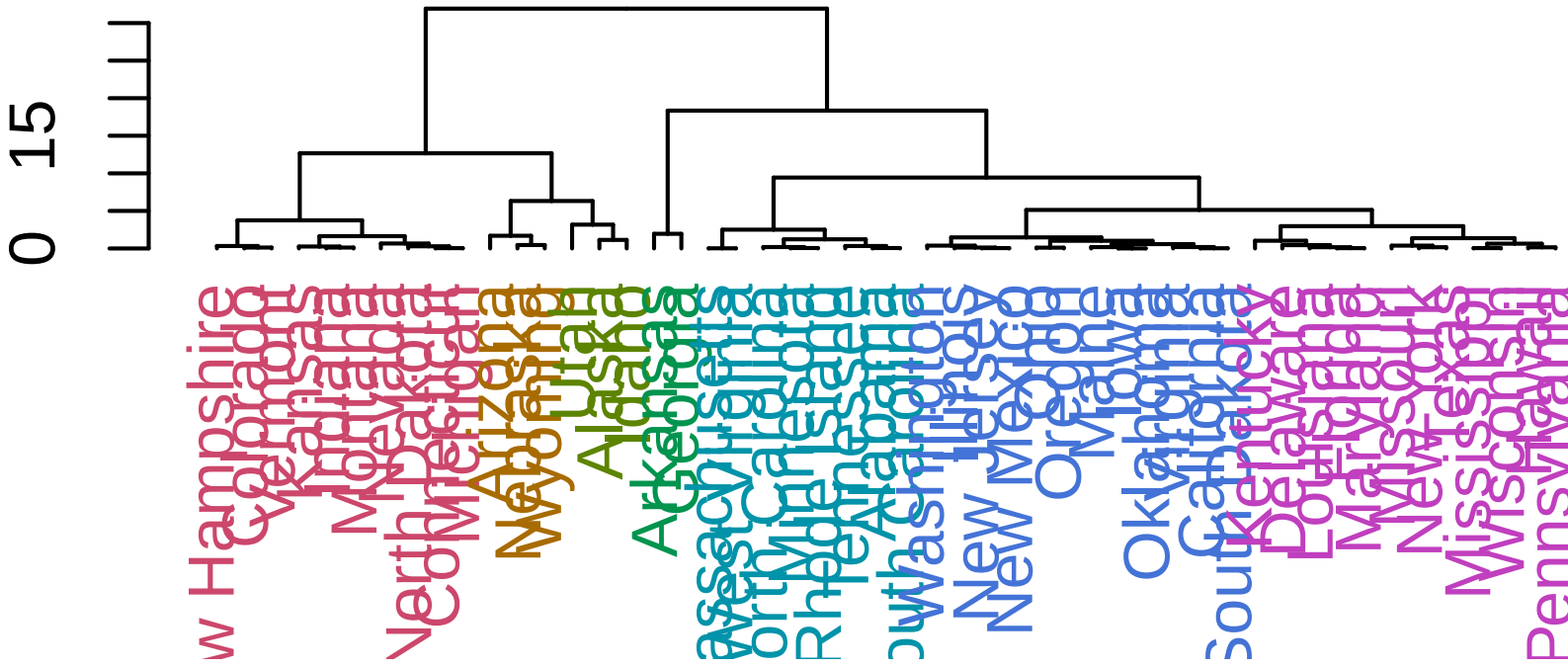
Machine Learning

Machine Learning

- Machine learning is a set of tools that aims to discover and predict data patterns.
- We can use machine learning to predict the outcome of a variable or do classification.
- You can also use machine learning on your data. Keep learning!
- Let us use Machine Learning to cluster voters in the 1976 Carter (Democrat) x Ford (Republican) election.
  - We rely on past Republican vote shares.


Machine Learning
Arizona
Alaska
California
Colorado
Connecticut
Idaho
Illinois
Indiana
Iowa
Kans
Mic
Wisconsin
West Virginia
Texas
Tennessee
South Carolina
Rhode Island
Pennsylva
North C
New
Miss
Florida
Delaware
Alabama
Arkansas
Wyoming
Washington
Virginia
Vermont
Utah
South Dakota
Oregon
Oklahoma
Ohio
North Dakota
Mexico
Jersey

Machine Learning



Shiny WebApps

Shiny WebApps

- Shiny WebApps are great for creating web applications to manipulate and visualize real-time data.
- We are not doing those here, but here are a few examples to explore. All done in R!
- Health Expenditure x Life Expectancy
- US Zipcode explorer
- Wordcloud creator

Shiny WebApps

- Bus company simulation game
- k-Means clustering
- COVID tracker
- Data analyzer
- You can also build your Shiny WebApp. Keep learning!

Scrollytells

Scrollytells

- Scrollytell is a nice wrapper for your Shiny. It helps you to present your analysis results using data.
- We are not doing those here, but here is an example for you to explore. All done in R!
- Scrollytell on Labor Automation
- You can also build your Scrollytell. Keep learning!

Questions?

Thank you all for a great quarter!