What can you do with R?

Vikram Dayal

Institute of Economic Growth

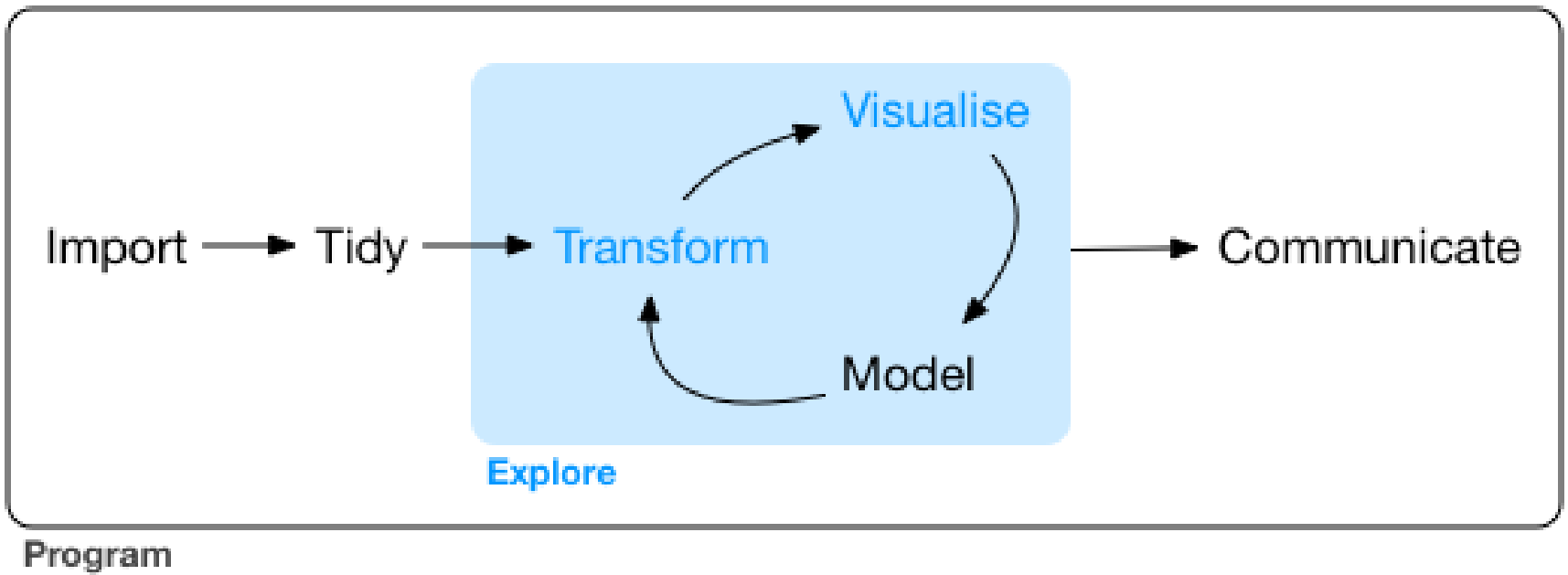
2021

The ability to take data —to be able to understand it, to process it, to extract value from it, to visualise it, to communicate it —that's going to be a hugely important skill ...

Because now we really do have essentially free and ubiquitous data. So the complementary factor is the ability to understand that data and extract value from it.'

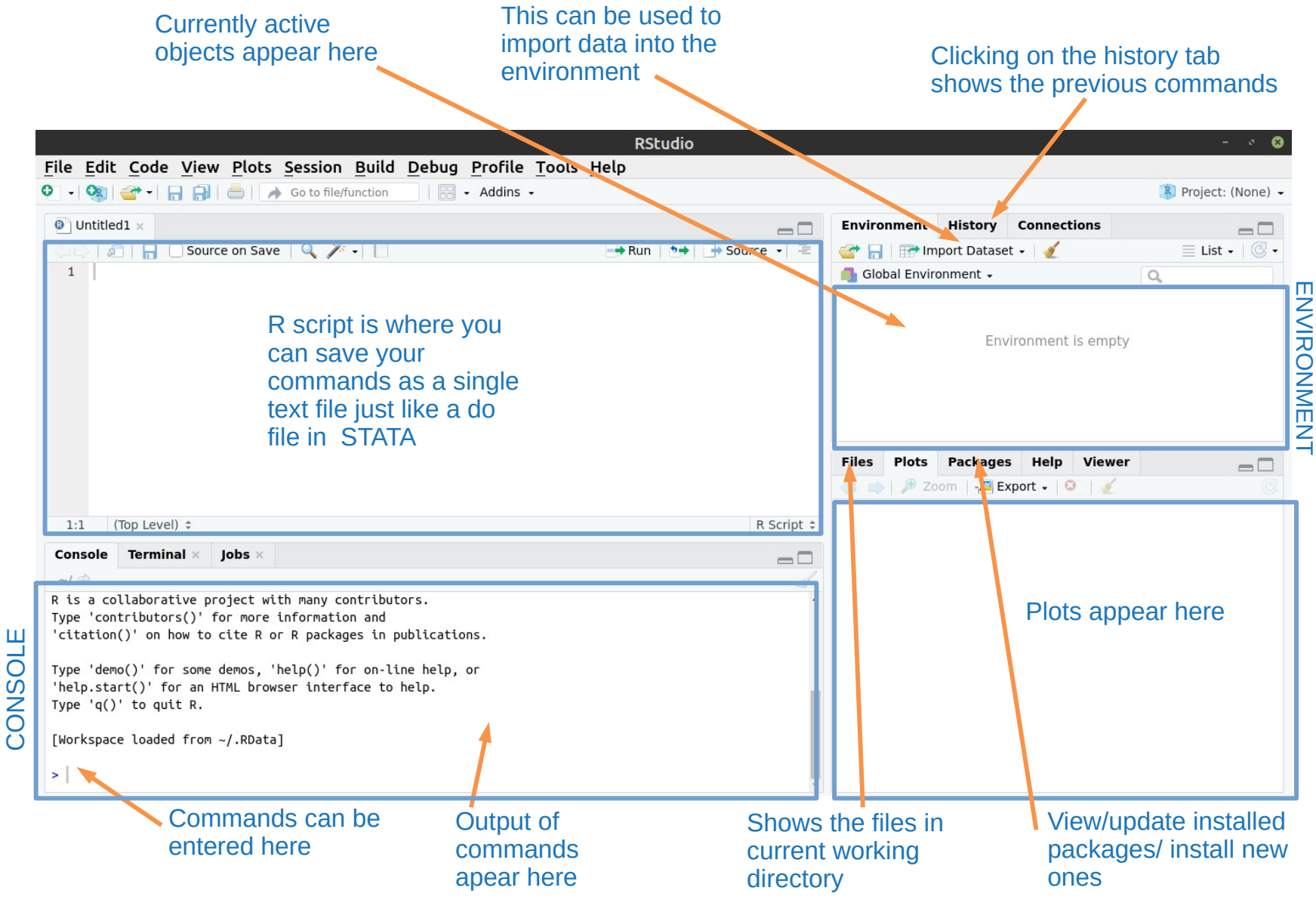
Hal Varian, Google Chief Economist, 2009

R is very good for the data analysis workflow



Hadley Wickham, author of the tidyverse

RStudio is a state of the art environment



Code records every step of your work

The Reinhart-Rogoff error – or how not to Excel at economics
April 23, 2013 6.40am AEST

https://theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646

... a famous 2010 academic paper, relied on by political big-hitters to bolster arguments for austerity cuts, contained significant errors.

... the errors in the Reinhart-Rogoff paper were not identified earlier can be ascribed by the pervasive failure of scientific and other researchers to make all data and computer code publicly available ...

Typical R Code

new object $\leftarrow$ function(object, information, options)

Example: create vector Price

Price $\leftarrow$ c(21,31,34)

The strength of R is the collection of packages

The *tidyverse* package is popular

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

You can do a lot with the tidyverse

```
library(tidyverse)
```

Get data into R

```
myfile <- read_csv("myfile.csv")
```

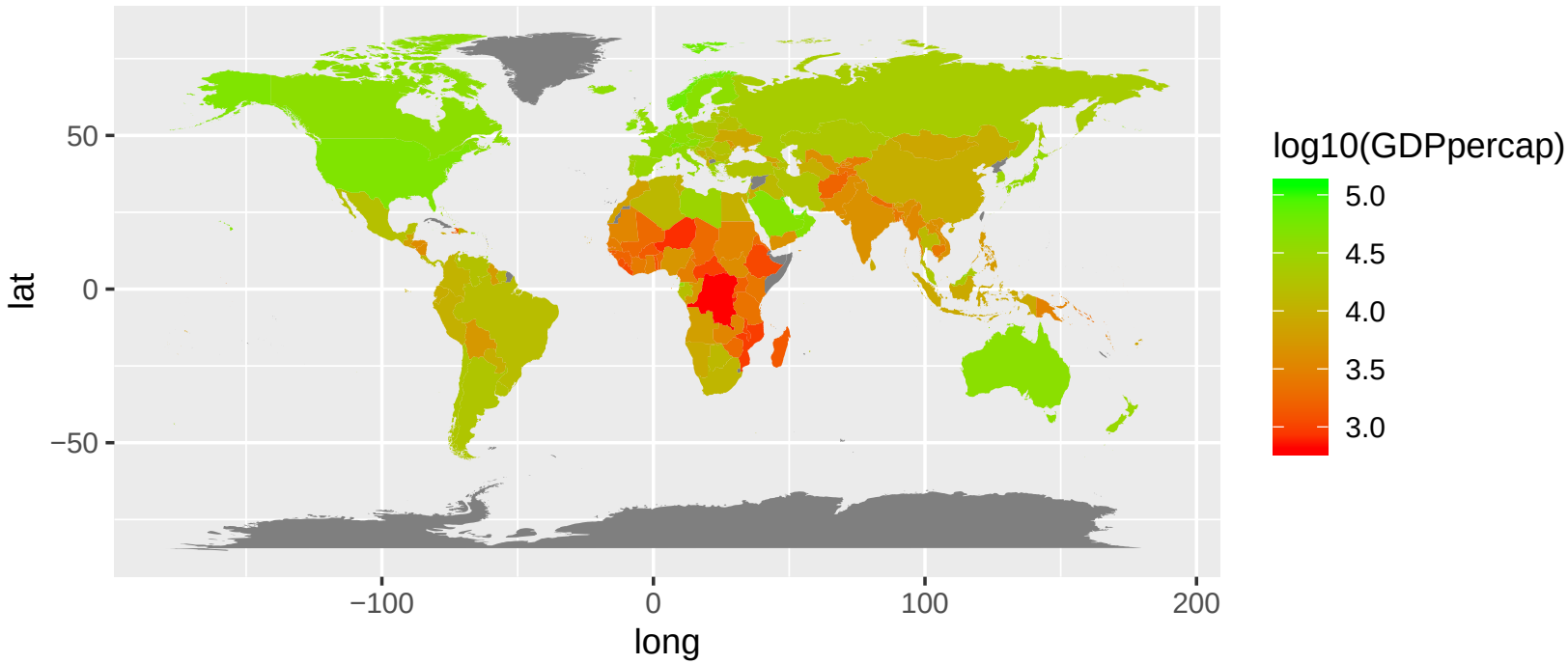
Plot

```
ggplot(mydata, aes(x = myx)) + geom_histogram()
```

Regression

```
reg.mod <- lm(y ~ x + z, data = mydata)
```

You can make maps with R



You can get the World Development Indicators data quickly with R

```
library(WDI)

WDIsearch("CO2.*capita")
```

You can work with networks in R

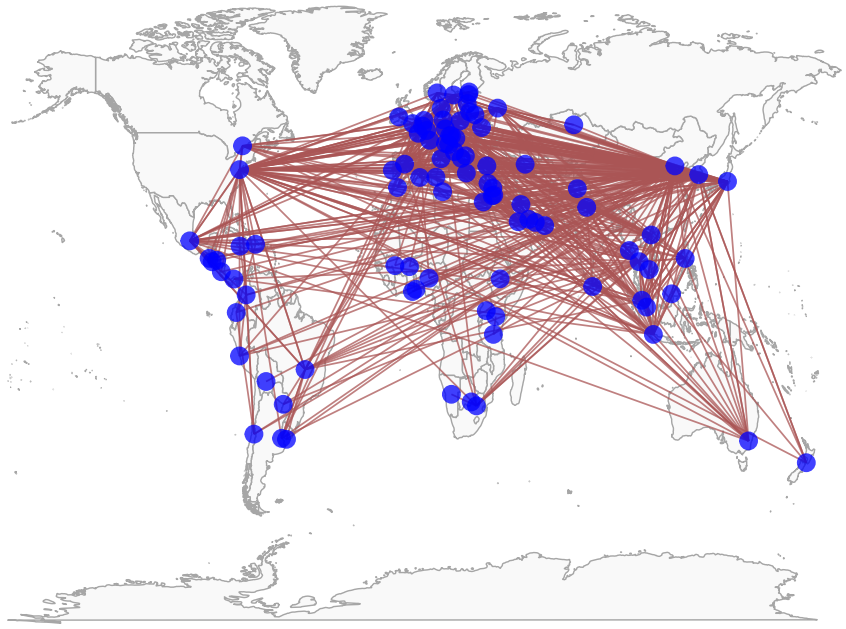
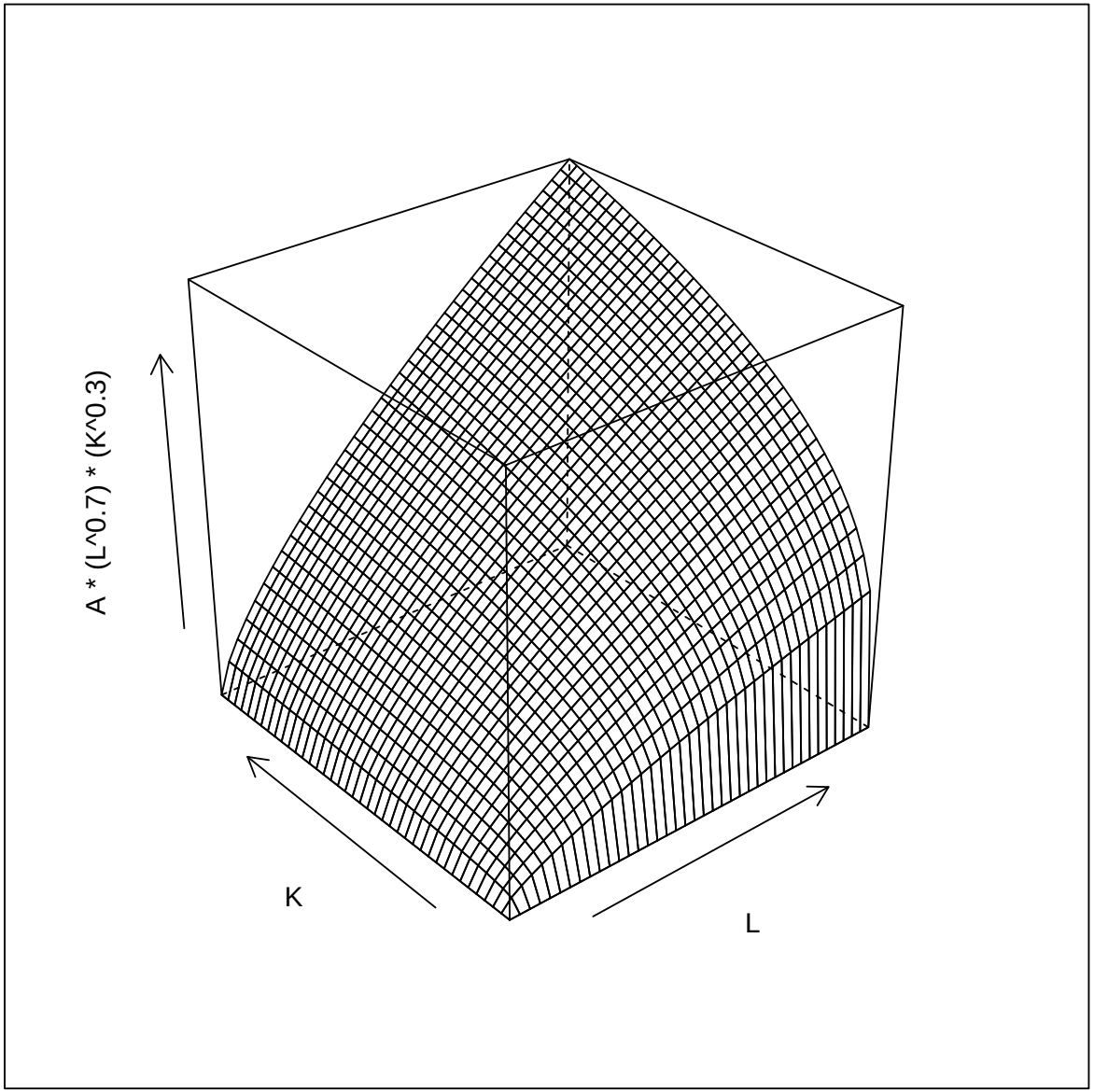


Figure: Electrical automotive goods 2016 network, ITNr package

You can plot mathematical functions with R



Causal inference

Most applied science is concerned with uncovering causal relationships. ...

The systematic use of RCTs to study causal relationships ... has resulted in tremendous welfare gains to society. ...

... many important questions ... researchers must rely on observational data ... But ... the underlying cause of any correlation remains unclear.

Scientific Background Nobel 2021

You can do randomization inference with R

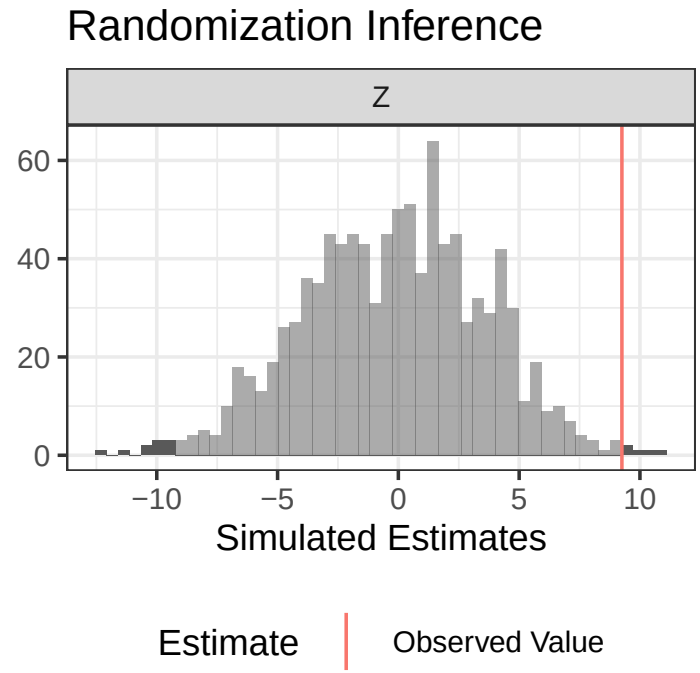


Figure: Duflo and Chattopadhyay: reserved villages had more new or repaired drinking water facilities in West Bengal

You can do regression discontinuity analysis with R

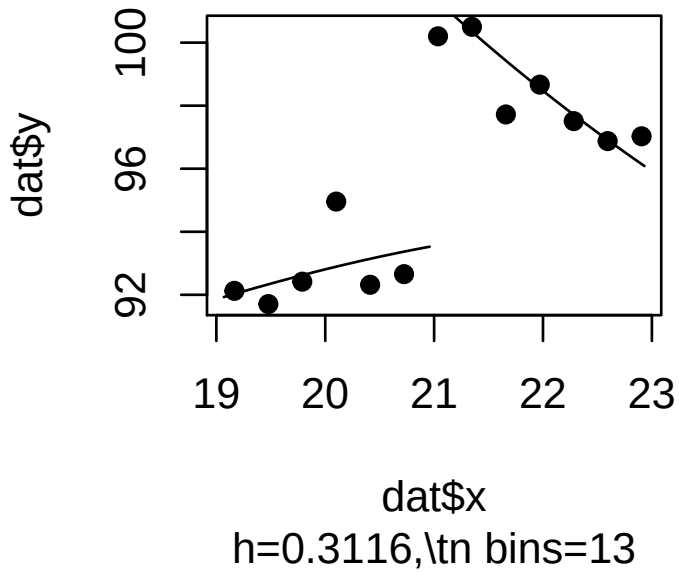


Figure: Did the Minimum Legal Drinking Age affect the death rate? (presented in Angrist and Pischke 2015), here using optimal bandwidth nonparametric regression (Imbens and coauthor)

You can do Monte Carlo simulations with R

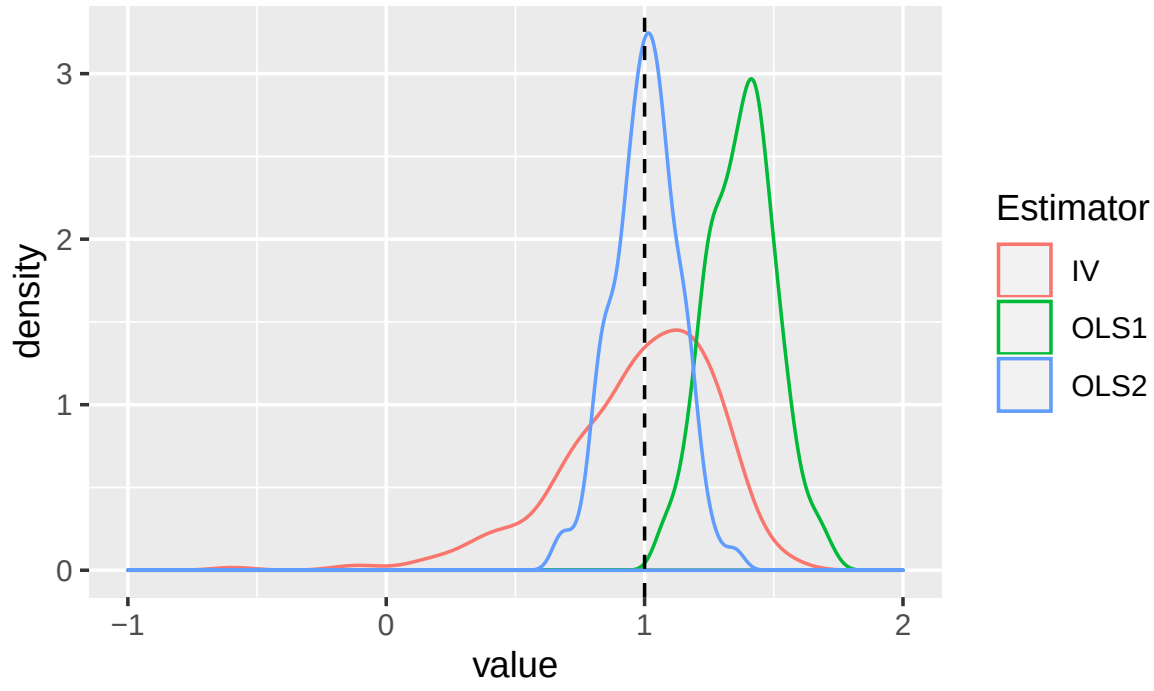


Figure: Sampling distributions of instrumental variables estimator (IV), OLS with omitted variable (OLS1) and OLS with 'correct' specification (OLS2).

Rather than use the theory to summarize the data through a set of structural parameters, it is sometimes more useful to present features of the data, often through ... graphical presentations ... and then to think about ... the process whereby they were generated.

Angus Deaton

You can plot covid cases with R

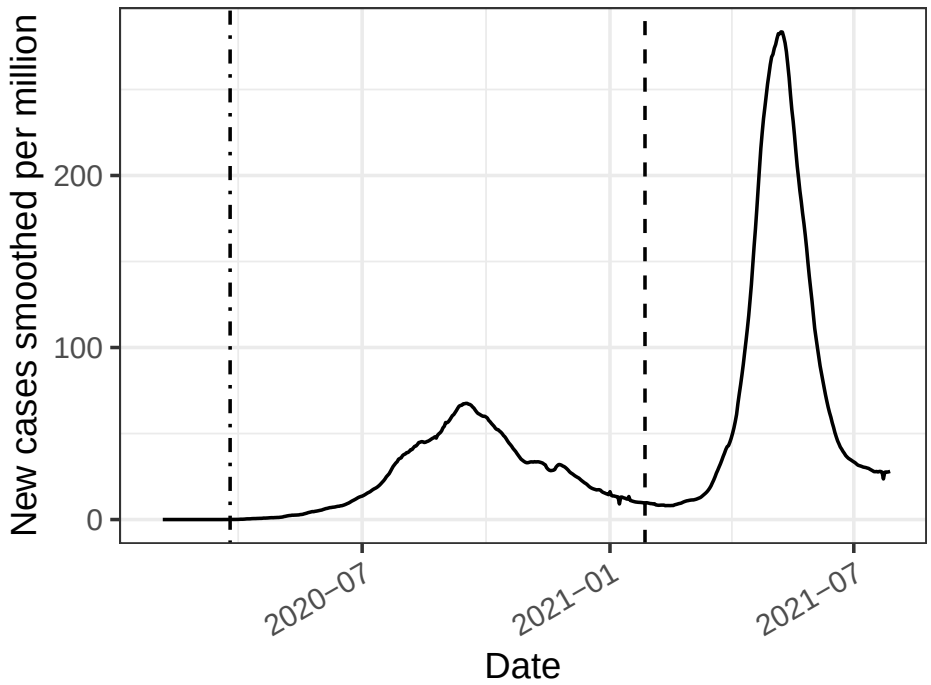


Figure: Covid cases in India

You can simulate the SIR epidemiological model with R

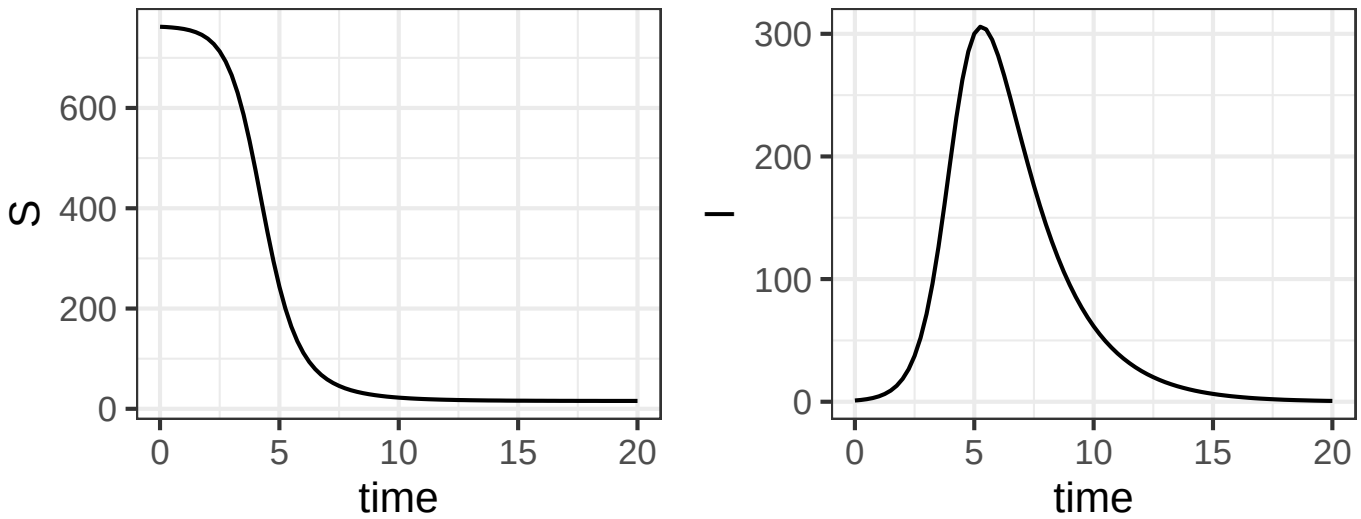


Figure: Simulated Susceptibles and Infections from SIR model

You can work with Google trends with R

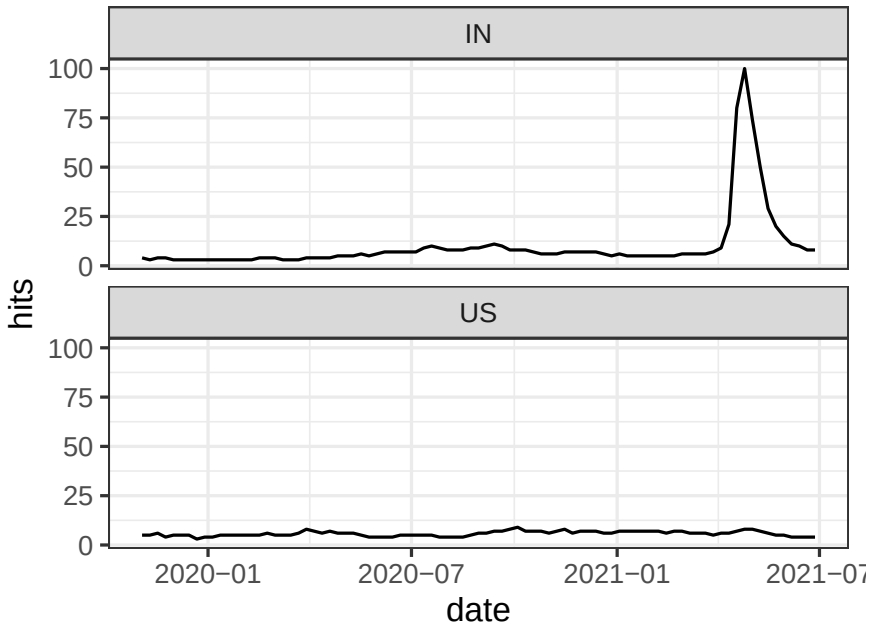


Figure: Google searches for ‘oxygen’

Predictive modeling ...

the focus of this type of modeling is to optimize prediction accuracy

... we don't really care why an email filter thinks a message is spam ... we only care that the filter accurately trashes spam

Kuhn and Johnson

You can use trees and random forests with R

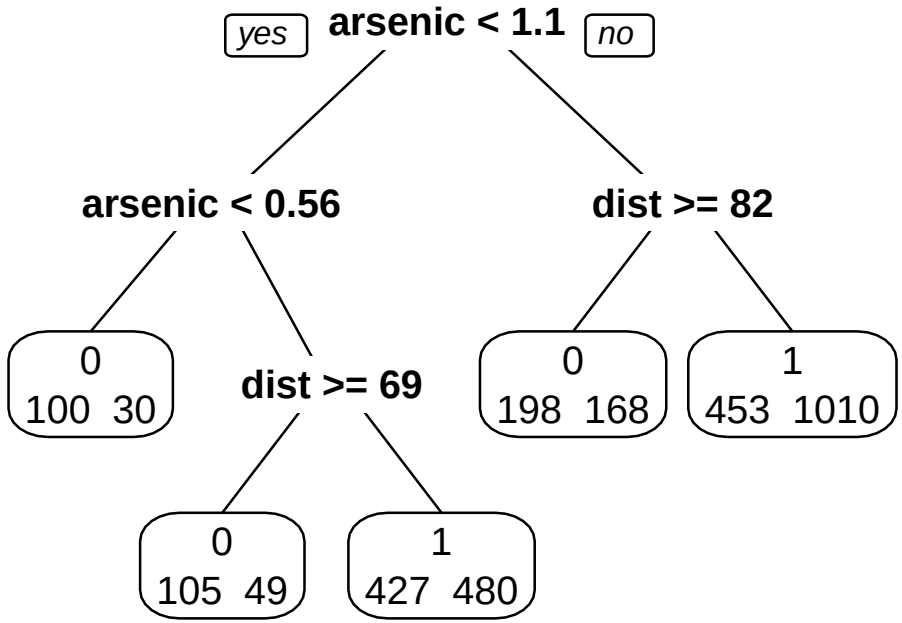


Figure: Classification tree for predicting well switching, in the context of arsenic in Bangladesh

You can communicate with RStudio

- You can combine text, code and output in one document or presentation
- You can make interactive web pages
- You can combine code from different software, and talk to different software

Thank you.