

ACCT 420: Course Logistics

Session 1

Dr. Richard M. Crowley

About Me

Teaching

- Third year at SMU
 - Previously taught ACCT 101
- Before SMU: Taught at the [University of Illinois Urbana-Champaign](#) while completing my PhD



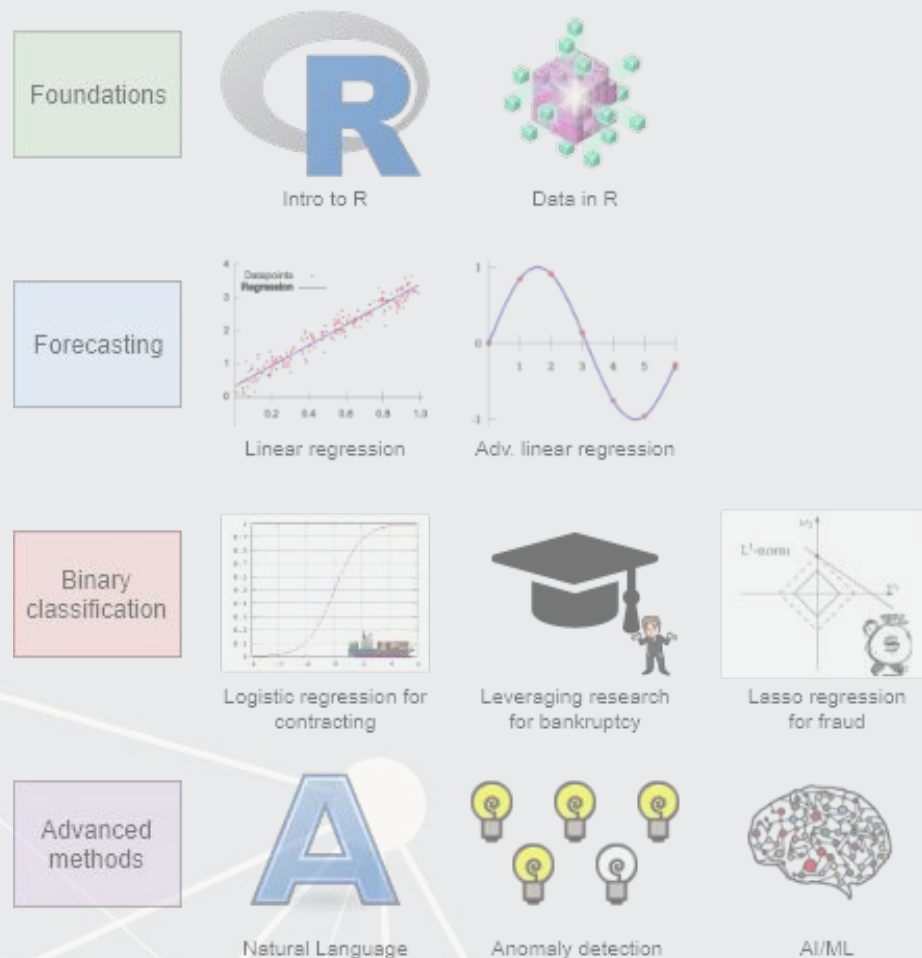
Research

- Accounting disclosure: What companies say, and why it matters



About this course

What will this course cover?



1. Foundations

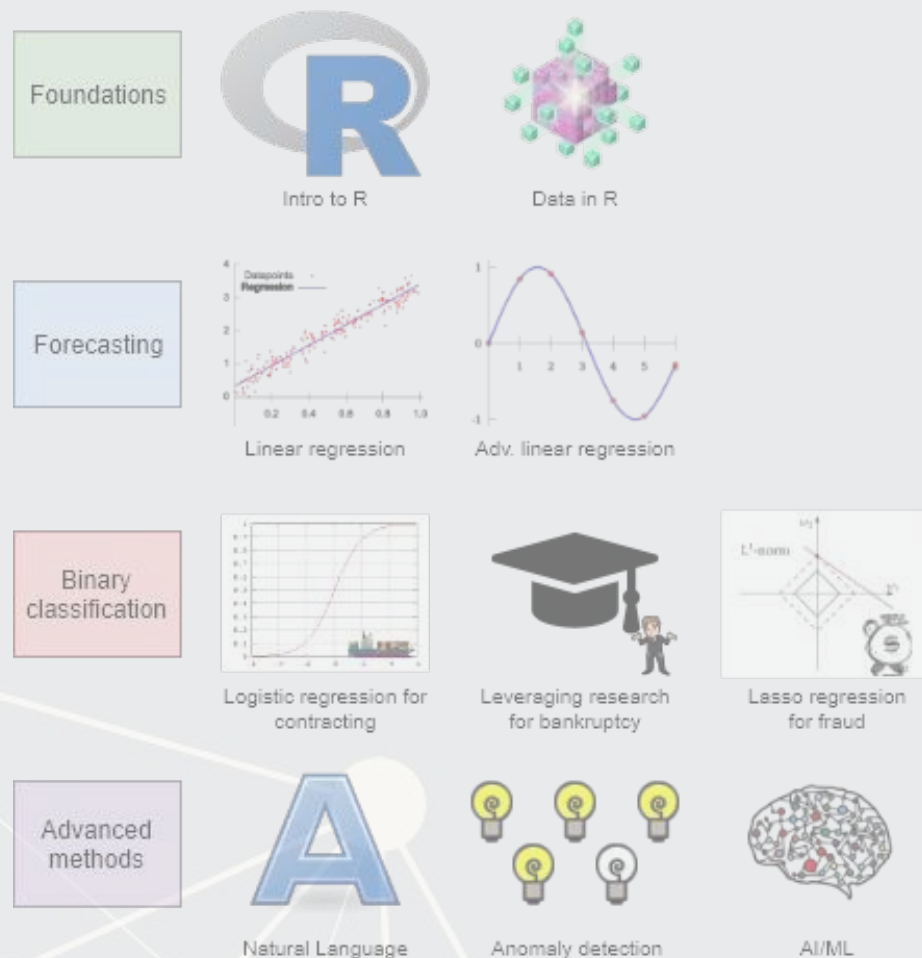
- Learning the ropes of R
- In class: Getting down the most important skills
- Outside: Practice and refining skills on Datacamp
- ~4 hours in week 1 and 2

2. Financial forecasting

- Predict financial outcomes
- Linear models

Learning and getting familiar with R and forecasting

What will this course cover?



3. Binary classification
 - Event prediction
 - Classification/detection
4. Advanced methods
 - Non-numeric data
 - Anomaly detection
 - AI/Machine learning
 - 2 weeks on current developments

Using R for higher level financial forecasting and detection

Datacamp

- Datacamp is providing *free* access to their *full* library of analytics and coding online tutorials
 - You will have free access for 6 months (Usually \$29 USD/mo)
- Online tutorials include short exercises and videos to help you learn R
- I have assigned materials via a Datacamp class, which will count towards participation
 - Check your email or eLearn for access
 - Datacamp automatically records when you finish these
 - I have personally done every assigned tutorial to verify their quality
- You are encouraged to go beyond the assigned materials – these will help you learn more about R and how to use it

Datacamp's tutorials teach R from the ground up, and are mandatory unless you can already code in R.

Textbook

- There is no required textbook
 - Datacamp is taking the place of the textbook
- If you prefer having a textbook...
 - [R for Everyone](#) by Jared Lander is a good one
- Other course materials (slides and articles) are available at:
 - eLearn
 - <https://rmc.link/acct420>
- Announcements will be only on Elearn

Teaching philosophy

1. Analytics is best learned by doing it
 - Less lecture, more thinking
2. Working with others greatly extends learning
 - If you are ahead:
 - The best sign that you've mastered a topic is if you can explain it to others
 - If you are lost:
 - Gives you a chance to get help the help you need

Grading

- Standard SMU grading policy
- Participation @ 10%
- Individual work @ 30%
- Group project @ 30%
- Final exam @ 30%



Participation

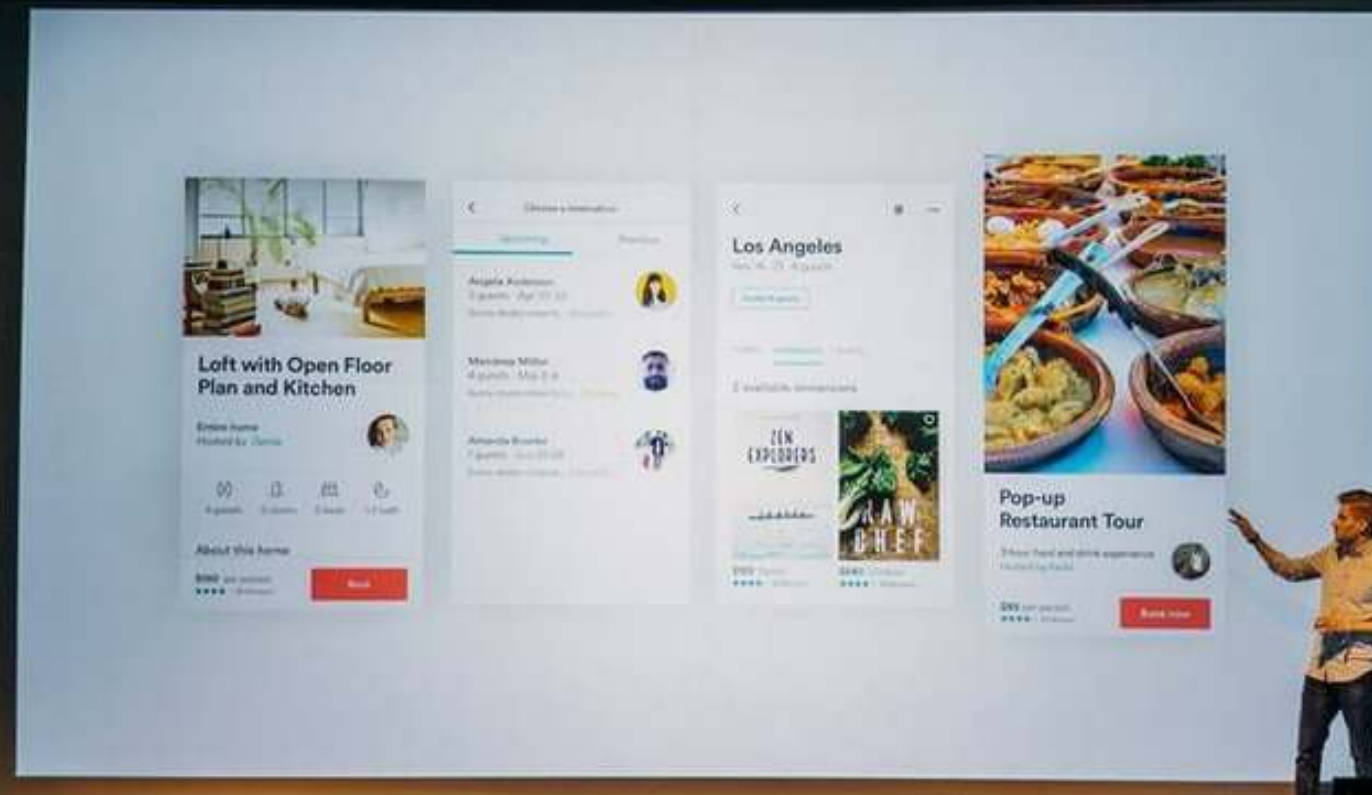
- Come to class
 - If you have a conflict, email me
 - Excused classes do not impact your participation grade
- Ask questions to **extend** or **clarify**
- Answer questions and explain answers
 - Give it your best shot!
- Help those in your group to understand concepts
- Present your work to the class
- Do the online exercises on Datacamp

Outside of class

- Verify your understanding of the material
- Apply to other real world data
 - Techniques and code will be useful after graduation
- Answers are expected to be your own work, unless otherwise stated
 - No sharing answers (unless otherwise stated)
- Submit on eLearn
- I will provide snippets of code to help you with trickier parts

Group project

To be announced later



Final exam

- Why?
 - Ex post indicator of attainment
- How?
 - Likely only 2 hours
 - Long format: problem solving oriented
 - Potentially a small amount of MCQ
- When?
 - Tentatively set for Tuesday, Dec 4 @ 1pm

Expectations

In class:

- Participate
 - Ask questions
 - Clarify
 - Add to the discussion
 - Answer questions
 - Work with classmates

Out of class

- Check eLearn for course announcements
- Do the assigned tutorials on Datacamp
 - This will make the course much easier!
- Do individual work on your own (unless otherwise stated)
 - Submit on eLearn
- Office hours are there to help!
 - Short questions can be emailed instead

Tech use

- Laptops and other tech are OK!
 - Use them for learning, not messaging
- Examples of good tech use:
 - Taking notes
 - Viewing slides
 - Working out problems
 - Group work
- Avoid:
 - Messaging your friends on Telegram
 - Working on homework for the class in a few hours
 - Watching livestreams of pandas or Hearthstone

Office hours

- Walk-in hours from 10:30-11:30am Fridays
 - Or by appointment
- Short questions can be emailed
 - I try to respond within 24 hours

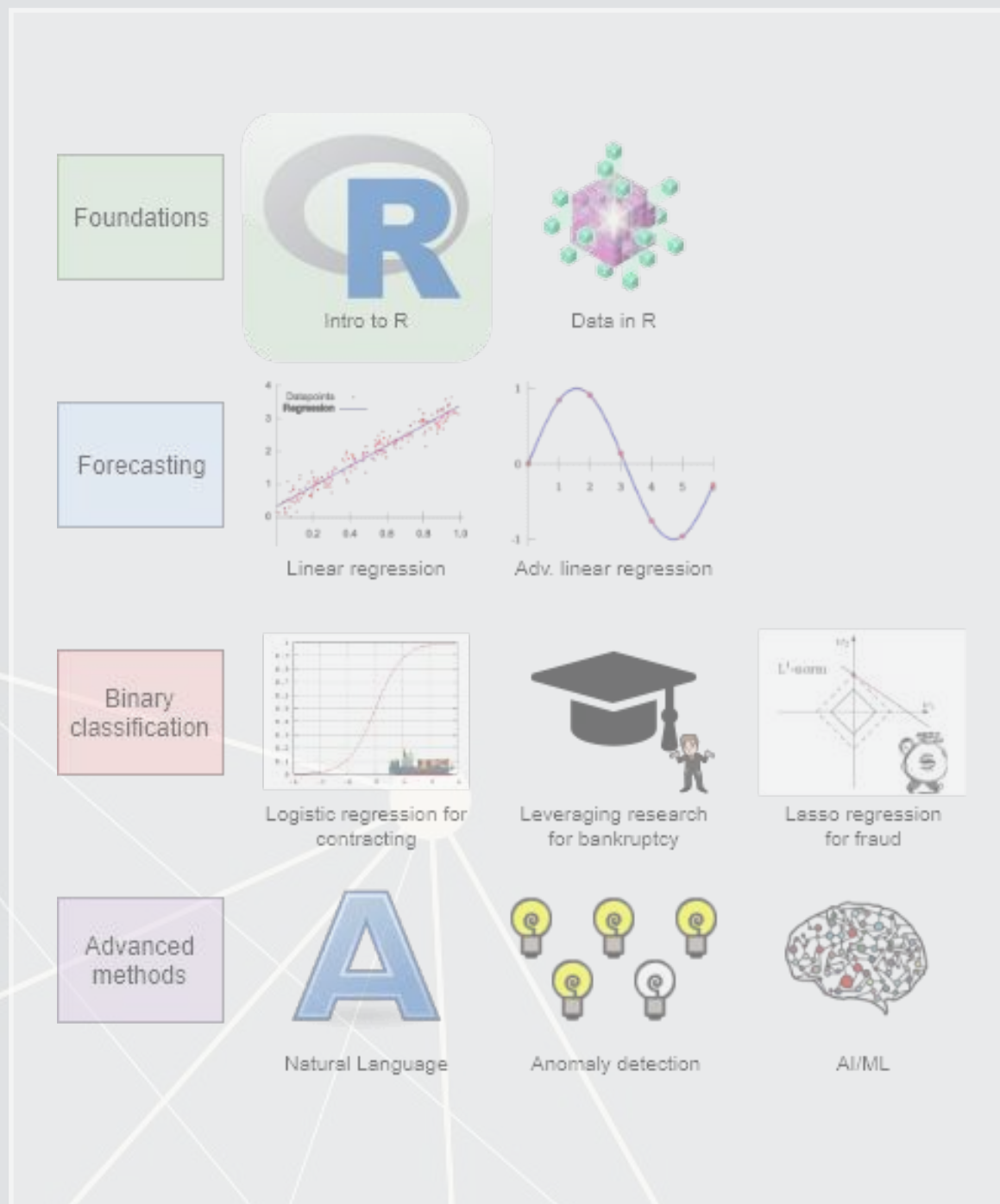
About you

About you

- Survey at rmc.link/aboutyou
- Results are anonymous
- We will go over the survey next week at the start of class

Introduction to analytics

Learning objectives



■ Theory:

- What is analytics?

■ Application:

- Who uses analytics? (and why?)

■ Methodology:

- Introduction to R

*Almost every class will touch on each of these three aspects

What is analytics?



What is analytics?

Oxford: The systematic computational analysis of data or statistics

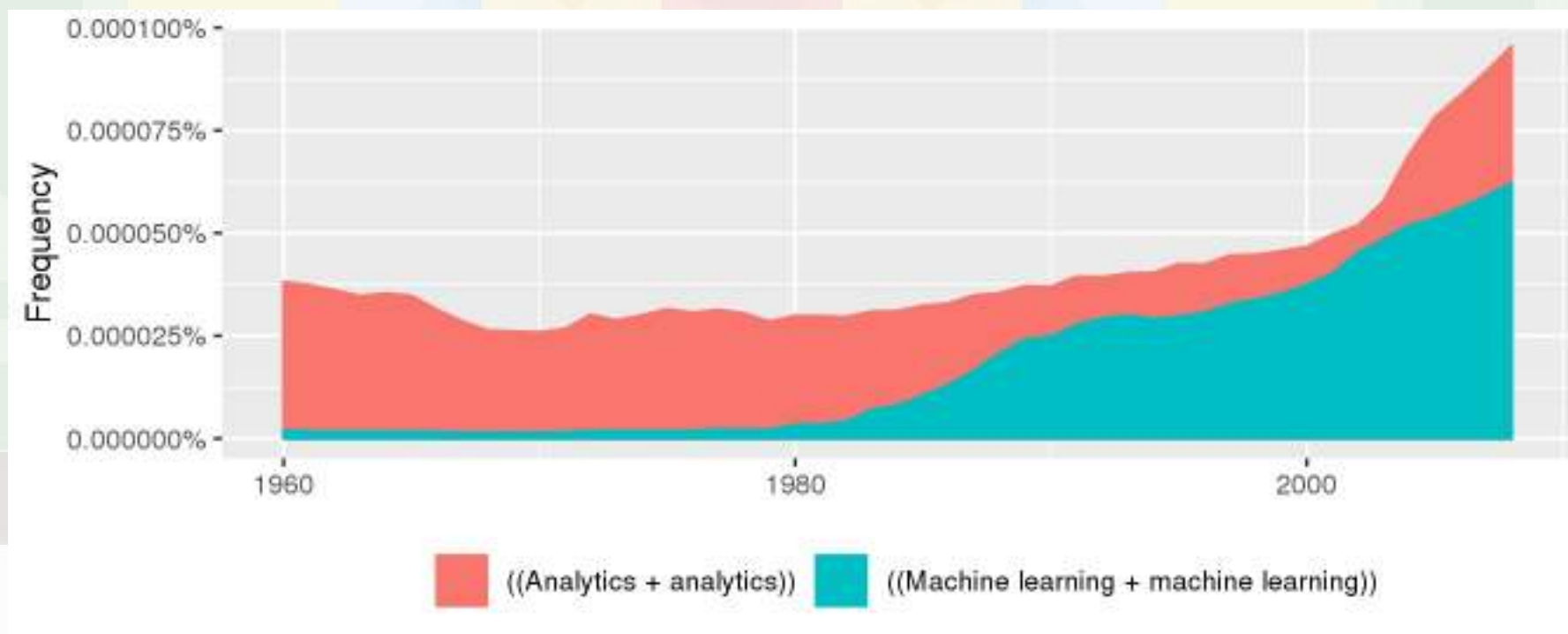
Webster: The method of logical analysis

Gartner: catch-all term for a variety of different business intelligence [...] and application-related initiatives

What is analytics?

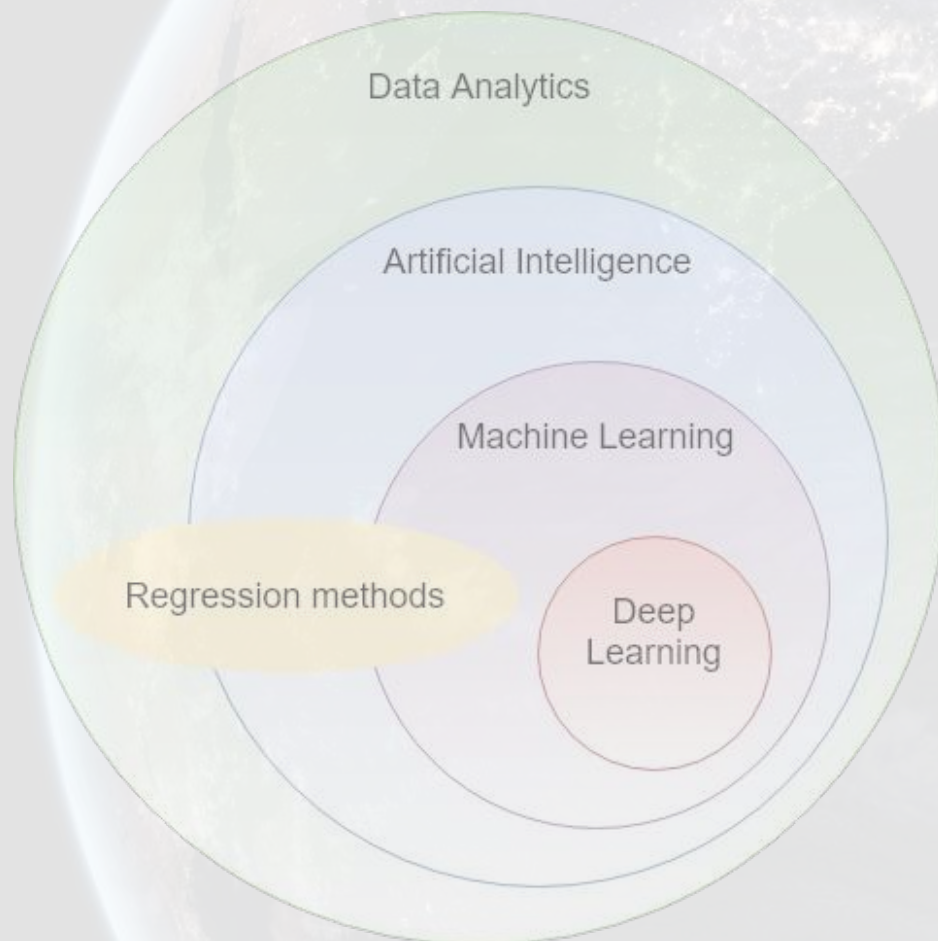
Simply put: Answering questions using data

- Additional layers we can add to the definition:
 - Answering questions using *a lot of* data
 - Answering questions using data *and statistics*
 - Answering questions using data *and computers*



Made using [seancarmody/ngramr](#)

Analytics vs AI/machine learning



- In class reading:
 - **AI Will Enhance Us, Not Replace Us**
 - By DataRobot's Senior Director of Product Marketing
 - Shortlink: rmc.link/420class1

How will Analytics/AI/ML change society and the accounting profession?

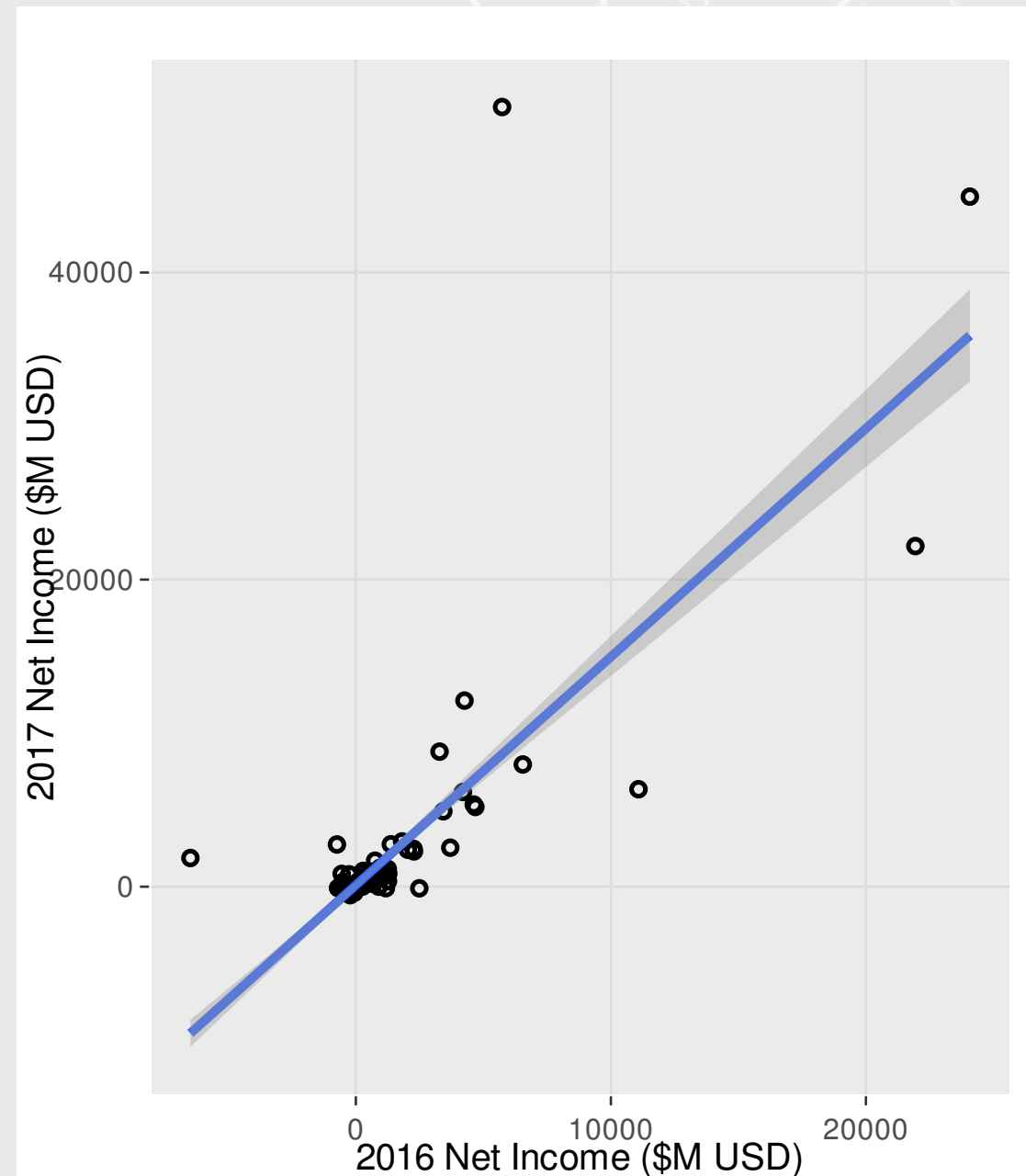
What are forecasting analytics?

- Forecasting is about making an educated guess of events to come in the future
 - Who will win the next soccer game?
 - What stock will have the best (risk-adjusted) performance?
 - What will Singtel's earnings be next quarter?
- Leverage *past* information
 - Implicitly assumes that the past and the future predictably related



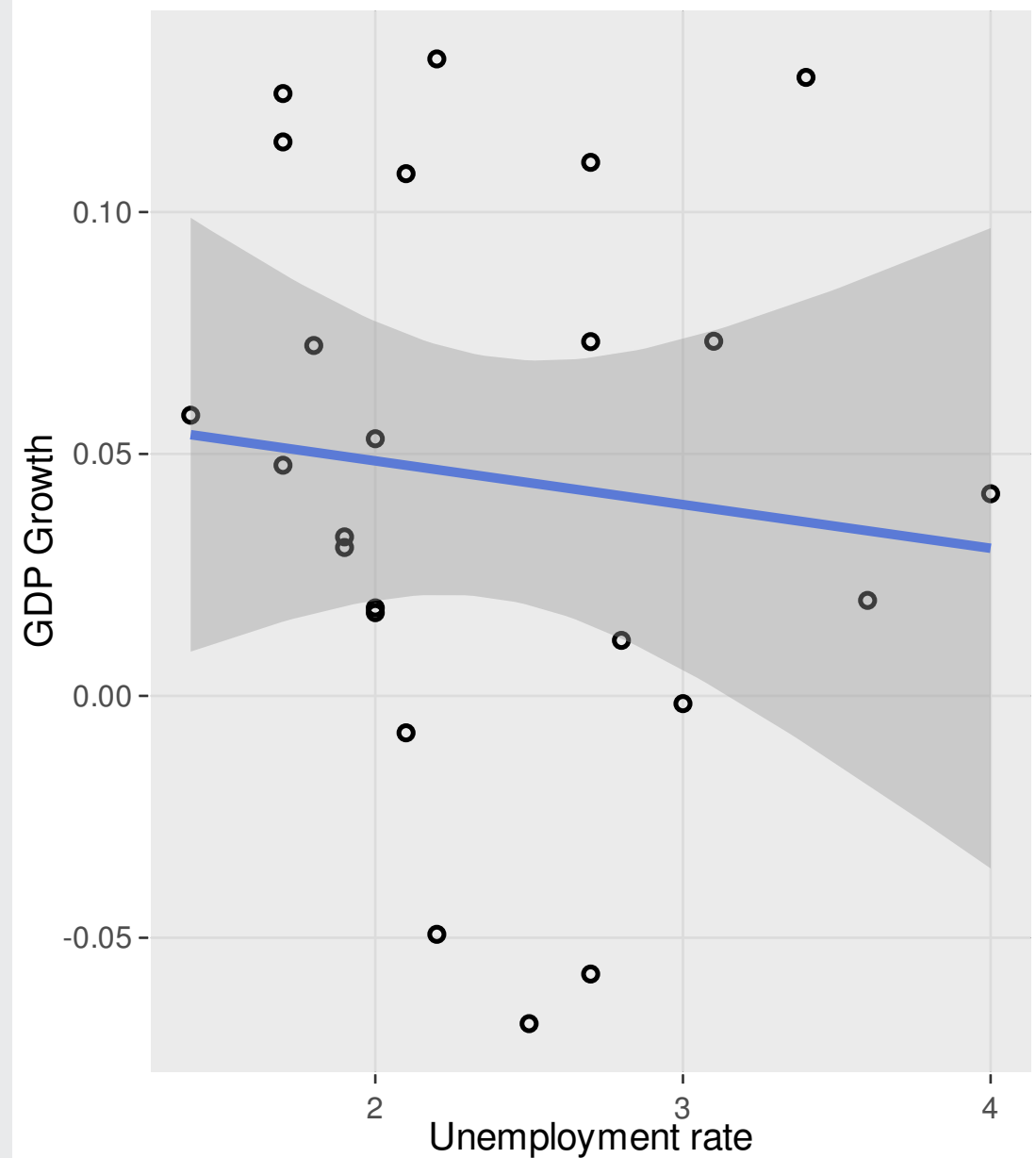
Past and future examples

- Past company earnings predicts future company earnings
- Some earnings are stable over time (Ohlsson model)
- Correlation: 0.7400142



Past and future examples

- Job reports predicts GDP growth in Singapore
- Economic relationship
- More unemployment in a year is related to lower GDP growth
- Correlation of -0.1047259

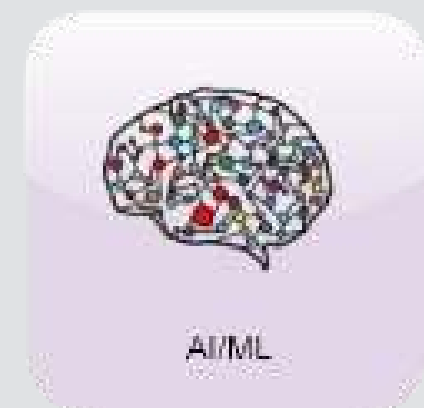
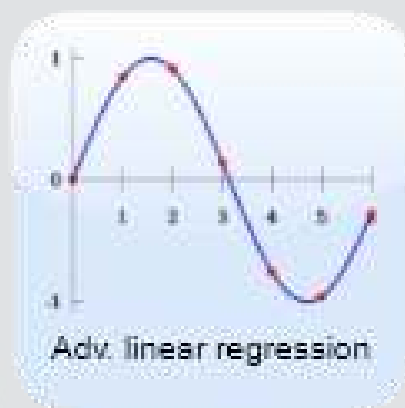
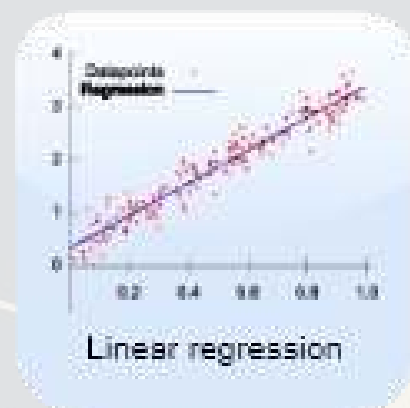


Past and future examples

- Ice cream revenue predicts pool drownings in the US
 - ???
 - Correlation is... only 0.0502886
- What about units sold?
 - Correlation is negative!!!
 - -0.720783
- What about price?
 - Correlation is 0.7872958

Forecasting analytics in this class

- Revenue/sales
- Shipping delays
- Bankruptcy
- Machine learning applications



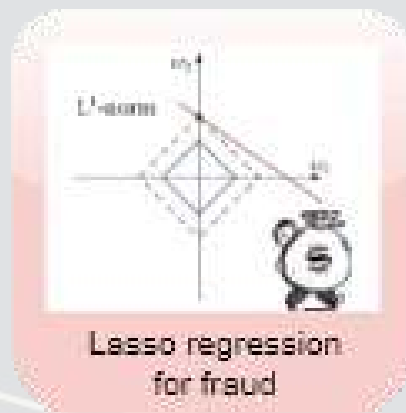
What are forensic analytics?

- Forensic analytics focus on *detection*
- Detecting crime such as bribery
- Detecting fraud within companies
- Looking at a lot of dog pictures to identify features unique to each breed



Forensic analytics in this class

- Fraud detection
- Working with textual data
- Detecting changes
- Machine learning applications



Forecasting vs forensic analytics

- Forecasting analytics requires a time dimension
 - Predicting *future* events
- Forensic analytics is about understanding or detect something
 - Doesn't need a time dimension, but it can help

These are not mutually exclusive. Forensic analytics can be used for forecasting!

Who uses analytics?

In general

- Governments
 - Al.Singapore
 - Big data office
 - “Smart” initiatives
- Academics
- Individuals!

- Companies
 - Finance
 - Manufacturing
 - Transportation
 - Computing
 - ...

53% of companies where using big data in a [2017 survey](#)!

What do companies use analytics for?

- Customer service
 - Royal Bank of Scotland
 - Understanding customer complaints
- Improving products
 - Siemens' Internet of Trains
 - Improving train reliability
- Their business
 - \$18.3B USD market in 2017
 - Just a small portion of overall IT spending (\$3.7T USD)



SIEMENS

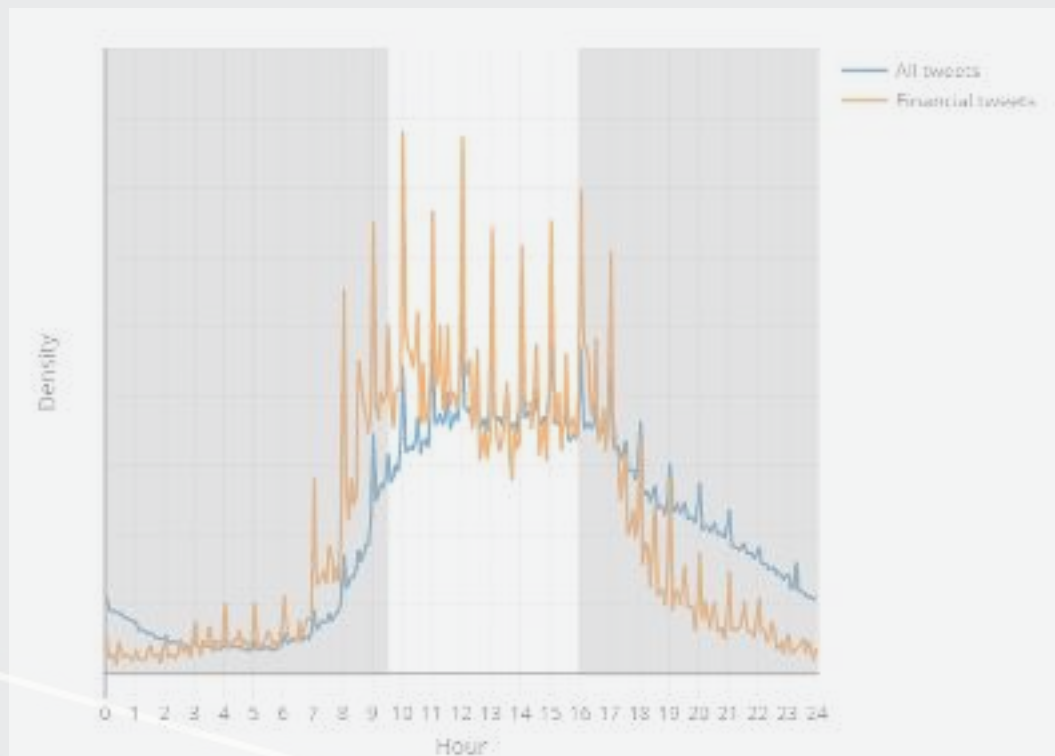
Gartner®

What do governments use analytics for?

- Govtech
 - Beeline
- Open data
 - Data.gov.sg
 - City of New York
- AI Singapore
 - Talent matching
 - 100 Experiments
 - AI in health Grand Challenge
 - AI research funding



What do academics use analytics for?



- Tweeting frequency by S&P 1500 companies ([paper](#))
- Aggregates every tweet from 2012 to 2016
- Shows frequency in 5 minute chunks
 - Note the spikes every hour!
- The white part is the time the NYSE is open

What do academics use analytics for?

- Annual report content that predicts fraud ([paper](#))
- For instance, discussing income is useful
 - first row is decreases,
 - second is increases
- But if it's good or bad depends on the year
- For instance, in 1999 it is a red flag
 - And one that Enron is flagged for



What do individuals use analytics for?



- Consulting
 - [Radim Řehůřek](#): Maintainer of [gensim](#), freelance consultant
- Investing
 - [Quantnet discussions](#)
- Health
 - Smart watches and other wearables

Why should you learn analytics?

- Important skill for understanding the world
 - Good timing to learn it, too!
- Gives you an edge over many others
 - Particularly useful for your career
- Jobs for “Management analysts” are expected to expand by 14% from 2016 to 2026
 - Accountants and auditors: 10%
 - Financial analysts: 11%
 - Average industry: 7%
 - All figures from US Bureau of Labor Statistics

Introduction to R

What is R?

- R is a “statistical programming language”
 - Focussed on data handling, calculation, data analysis, and visualization
- We will use R for all work in this course



Why do we need R?

- Analytics deals with more data than we can process by hand
 - We need to ask a computer to do the work!
- R is one of the de facto standards for analytics work
 - Third most popular language for data analytics and machine learning ([source](#))
 - Fastest growing of all mainstream languages
 - Free and open source, so you can use it anywhere
 - It can do most any analytics
 - Not a general programming language

Programming in R provides a way of talking with the computer to make it do what you want it to do

Setup

- For this class, I will assume you are using RStudio with the default R installation
 - RStudio downloads
 - R for Windows
 - R for (Max) OS X (Download R-3.5.1.pkg)
 - R for Linux
- You will need a laptop or desktop for this
 - I am working to find a lab on campus for this as well
- For the most part, everything will work the same across all computer types
- Everything in these slides was tested on R 3.5.0 and 3.5.1

How to use R Studio

1. R markdown file

- You can write out reports with embedded analytics

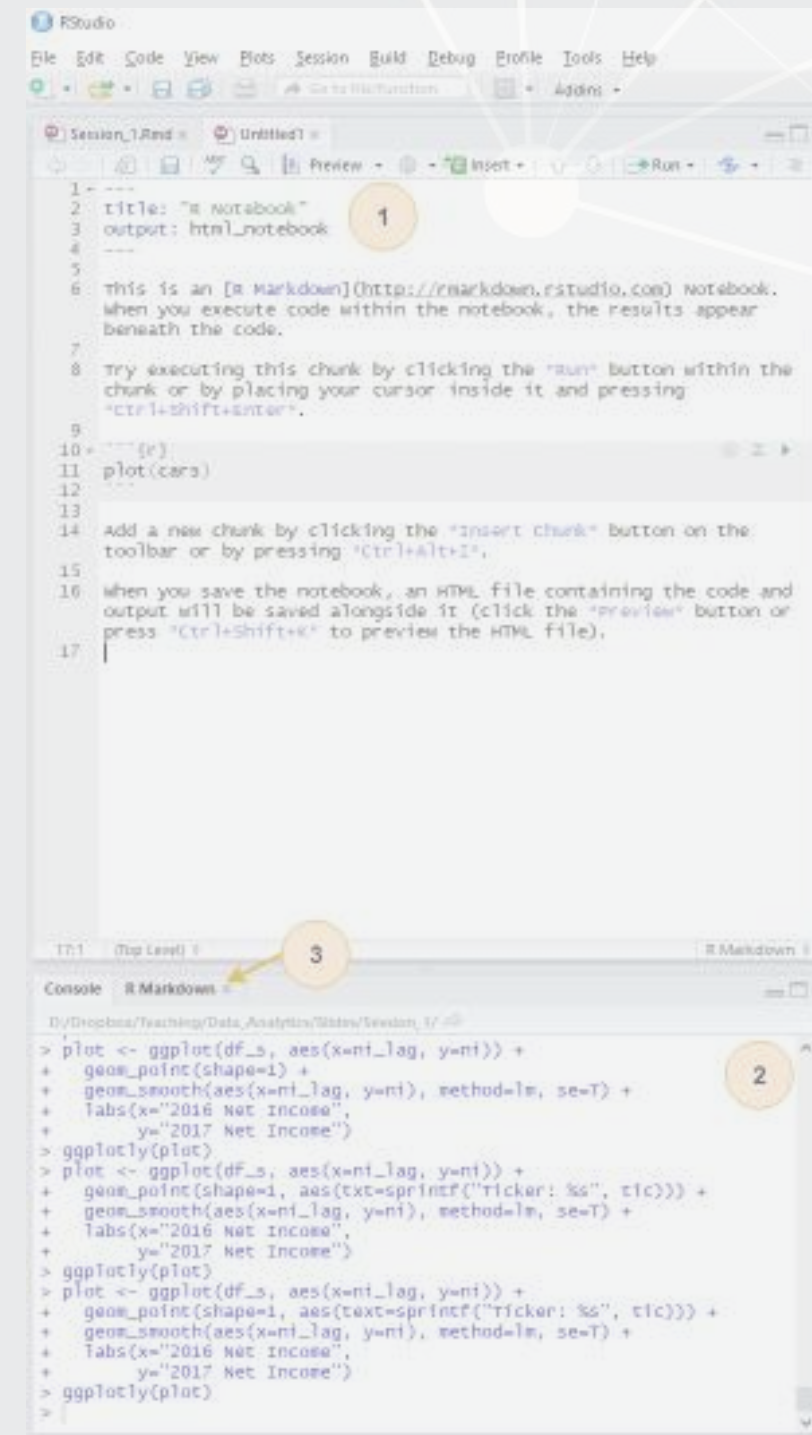
2. Console

- Useful for testing code and exploring your data

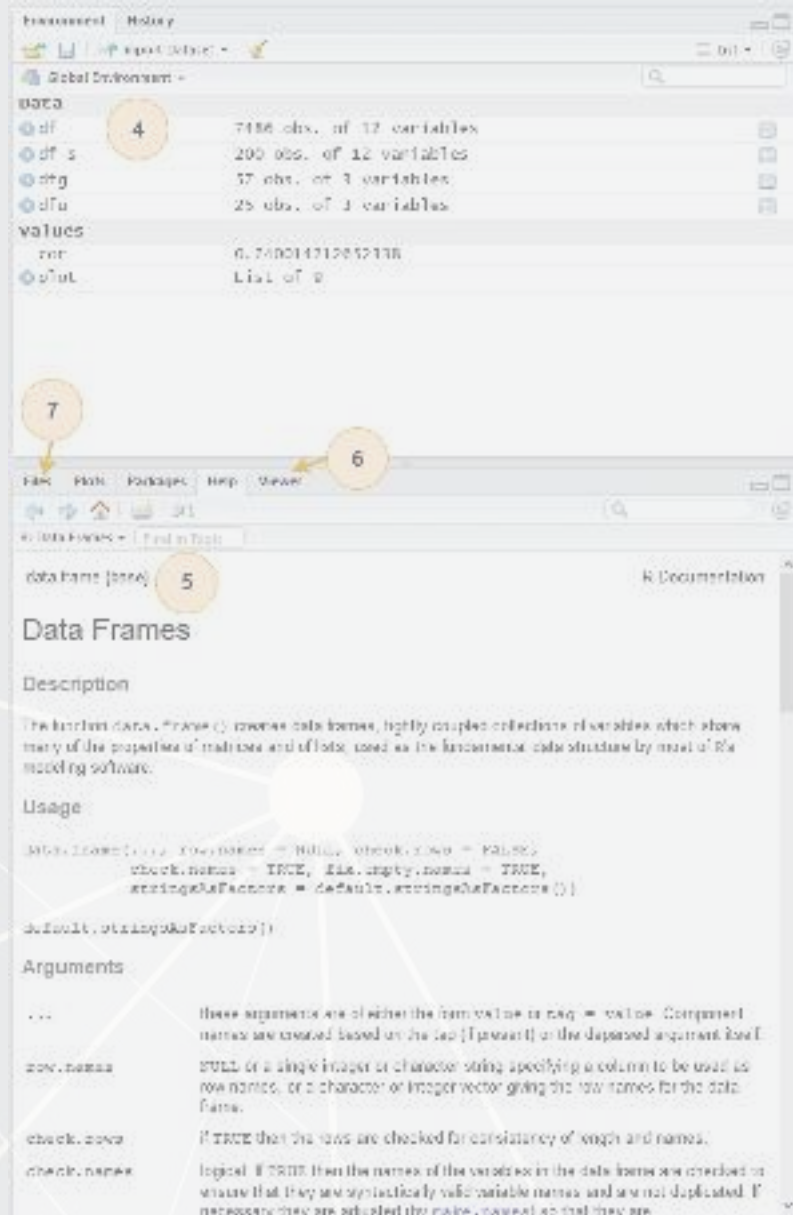
- Enter your code one line at a time

3. R Markdown console

- Shows if there are any errors when preparing your report



How to use R Studio



4. Environment

- Shows all the values you have stored

5. Help

- Can search documentation for instructions on how to use a function

6. Viewer

- Shows any output you have at the moment.

7. Files

- Shows files on your computer

Basic R commands

Arithmetic

- Anything in boxes like those on the right in my slides are R code
- The slides themselves are made in R, so you could copy and paste any code in the slides right into R to use it yourself
- Grey boxes: Code
 - Lines starting with `#` are comments
 - They only explain what the code does
- Blue boxes: Output

```
# Addition uses '+'  
1 + 1
```

```
## [1] 2
```

```
# Subtraction uses '-'  
2 - 1
```

```
## [1] 1
```

```
# Multiplication uses '*'  
3 * 3
```

```
## [1] 9
```

```
# Division uses '/'  
4 / 2
```

```
## [1] 2
```

Arithmetic

- Exponentiation
 - Write x^y as `x ^ y`
- Modulus
 - The remainder after division
 - Ex.: $46 \bmod 6 = 4$
 1. $6 \times 7 = 42$
 2. $46 - 42 = 4$
 3. $4 < 6$, so 4 is the remainder
- Integer division (not used often)
 - Like division, but it drops any decimal

Exponentiation uses '^'

`5 ^ 5`

`## [1] 3125`

Modulus (aka the remainder) uses '%%'

`46 %% 6`

`## [1] 4`

Integer division uses '%/%'

`46 %/% 6`

`## [1] 7`

Variable assignment

- Variable assignment lets you give something a name
 - This lets you easily reuse it
- In R, we can name almost anything that we create
 - Values
 - Data
 - Functions
 - etc...
- We will name things using the `<-` command

```
# Store 2 in 'x'
```

```
x <- 2
```

```
# Check the value of x
```

```
x
```

```
## [1] 2
```

```
# Store arithmetic in y
```

```
y <- x * 2
```

```
# Check the value of y
```

```
y
```

```
## [1] 4
```

Variable assignment

- Note that values are calculated at the time of assignment
- We previously set `y <- 2 * x`
- If we change the values of `x` and `y` remain unchanged!

```
# Previous value of x and y  
x
```

```
## [1] 2
```

```
y
```

```
## [1] 4
```

```
# Change x, then recheck the value  
# of x and y  
x <- 200  
  
x
```

```
## [1] 200
```

```
y
```

```
## [1] 4
```

Application: Singtel's earnings growth

Set a variable growth to the amount of Singtel's earnings growth percent in 2018

```
# Data from Singtel's earnings reports, in Millions of SGD
```

```
singtel_2017 <- 3831.0
```

```
singtel_2018 <- 5430.3
```

```
# Compute growth
```

```
growth <- singtel_2018 / singtel_2017 - 1
```

```
# Check the value of growth
```

```
growth
```

```
## [1] 0.4174628
```

Recap

- So far, we are using R as a glorified calculator
- The key to using R is that we can scale this up with little effort
 - Calculating *every* public companies' earnings growth isn't much harder than calculating Singtel's!

Scaling this up will give use a lot more value

- We can also leverage **functions** to automate more complex operations
 - There are many functions built in, and many more freely available
 - *We'll cover this next week*
- We'll also need ways to read **data files** and work with collections of numbers
 - *We'll cover this next week as well*

Wrap up

- R Practice
 - Shortlink: rmc.link/420r1
 - Do the practice here if you would like help with it
 - Otherwise, do it at home
- For next week:
 - Start working on the Datacamp tutorials!
 - Assigned tutorials are on the Datacamp class page
 - For next week, complete the *Intro to R* course
 - More tutorials will be assigned in future weeks
 - Other helpful tutorials:
 - [Rmarkdown tutorial from RStudio](#)