Week One

Introduction to Data Science

Philip Leftwich

27.9.2021

COVID-safety

- Information UEA COVID-19 Advice
- Take Lateral flow tests on **Mondays** & **Thursdays**
- Self-isolate and get a **PCR** test if unwell or any symptoms
- Wear **Face-coverings** indoors

Hello! 👋

Welcome to Data Science

- How are you doing today?
- What made you sign up for this module?
- Go to Slido.com #602443

Attendance

8t8s99

BIO-5023YA21002 - Fri 01 Oct 21

Timetables

- Check *Timetabler* regularly for updates/changes
- One lecture per week - In-person/Collaborate
- One workshop per week
  - Can bring own laptop
  - Streamed but not recorded

Blackboard

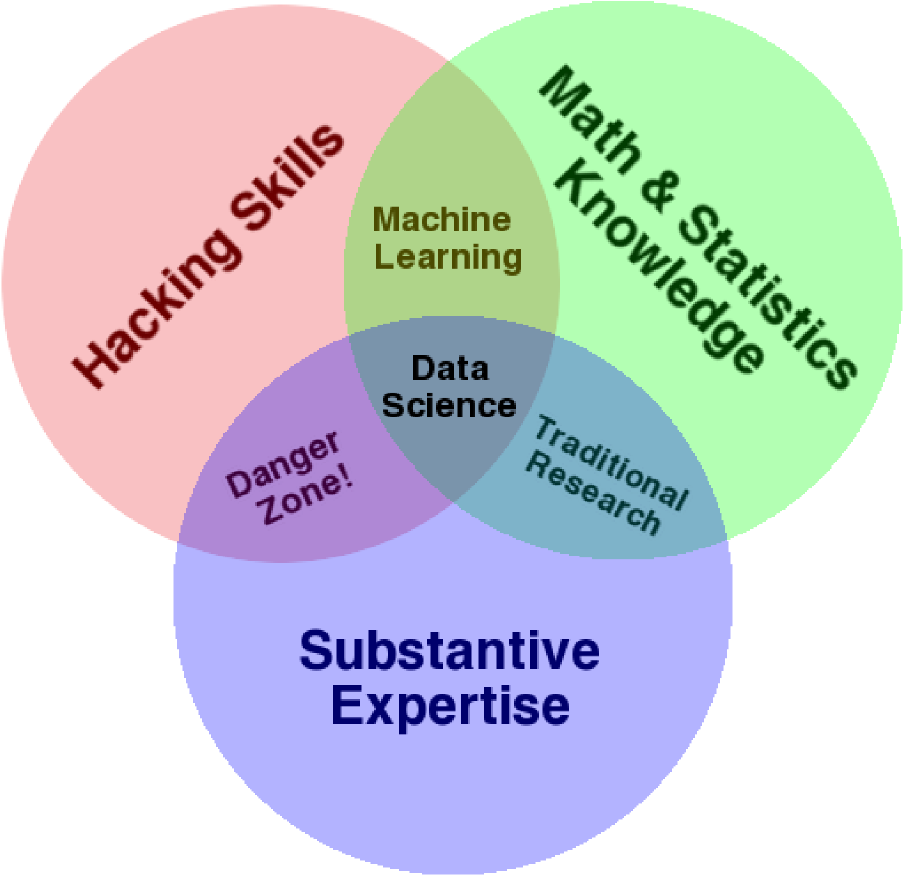
- Announcements
- Yammer Feed
- Lecture Slides
- Module Information
- Collaborate Link
- Assessment Briefs

Assessment

All coursework, no exam

- 40% Summative this term
- 60% Summative next term

What is Data Science?



Insights from Data

- Clear, deep understanding of a complicated problem or situation

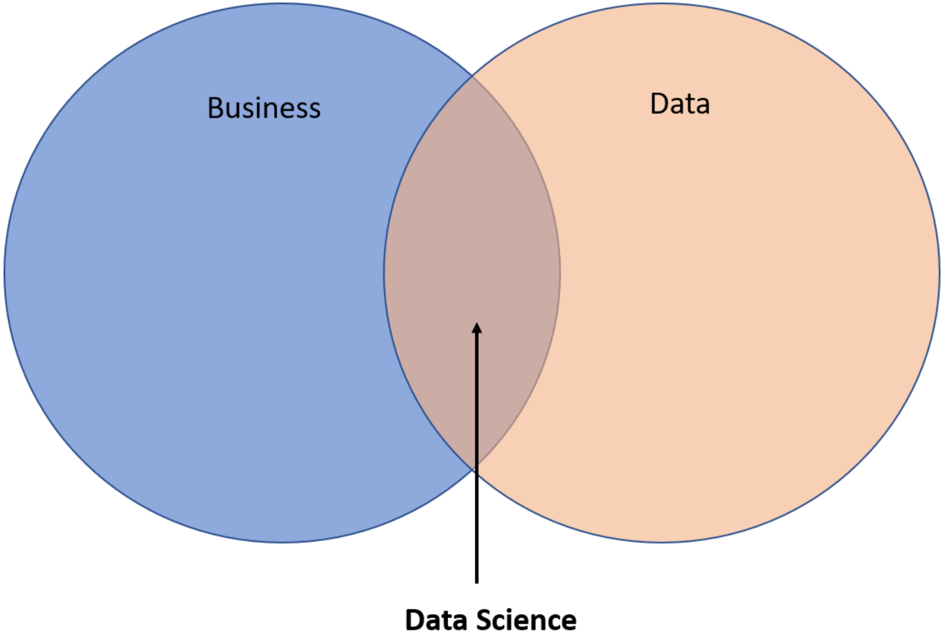
Insights from Data

- Clear, deep understanding of a complicated problem or situation
- Become better scientists

Insights from Data

- Clear, deep understanding of a complicated problem or situation
- Become better scientists
- Gain programming and analysis skills that are in demand by business

Data is big business



Data is big business

- Fitbit
- Amazon
- Aviva
- Open Health Foundation
- Local/National COVID strategies

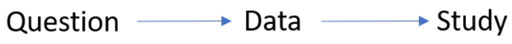
What can Data do?

- Increase Revenues
- Open new markets
- Improve efficiencies
- Provide insights
- Make predictions

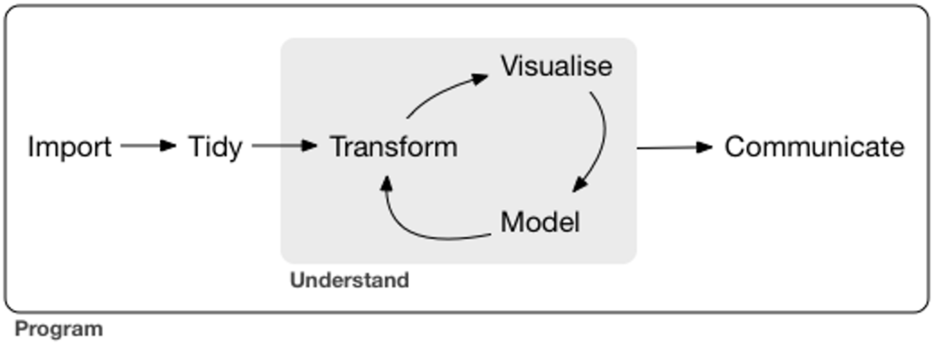
Become a better scientist

- A better platfom to generate and test hypotheses
- Produce better data visuals
- Gain the statistical tools to describe and predict from data
- Understand the importance of "open" and "reproducible" research

Our process



Question → Data → Study



- We will use the programming language R - it is fun, flexible and will empower you to be a better Scientist.

Questions

The starting point of gaining insights should always be the Question, not the Data

- Is there a pattern/relationship that matches our expectations?
- Can we ascribe causation?
- Can we make predictions?

Hypothesis

Turn a question into a hypothesis

Where does data come from?

- Controlled experiments
- 'Field' experiments
- Exploratory studies

Data

What is data?

- Data are records/observations/measurements

Data

What is data?

- Data are records/observations/measurements

- Data can be quantitative

  - Values
  - Continuous
  - Integer
  - Categorical

Data

What is data?

- Data are records/observations/measurements
- Data can be quantitative
    - Values
    - Continuous
    - Integer
    - Categorical
- Data can be qualitative
    - Opinion polls
    - Text mining
    - Colours

Insights from Data

- A dataset is a collection of data
- There are many ways to arrange datasets
- We aim to cut through the variation/noise to identify patterns

Key features of datasets

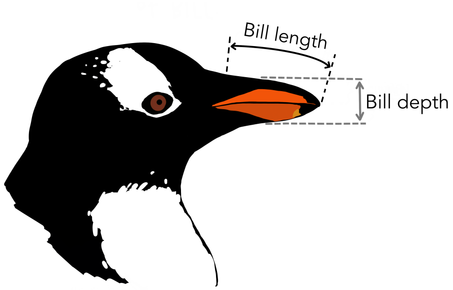
- Observations
- Variables
  - Response
  - Predictor
- Correlations among variables/ Confounding effects
- Independence of observations

Example of Insights

This is from the palmer penguins dataset curated by Dr. Allison Horst. Data were originally collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER.

Question What is the relationship between bill length and bill depth

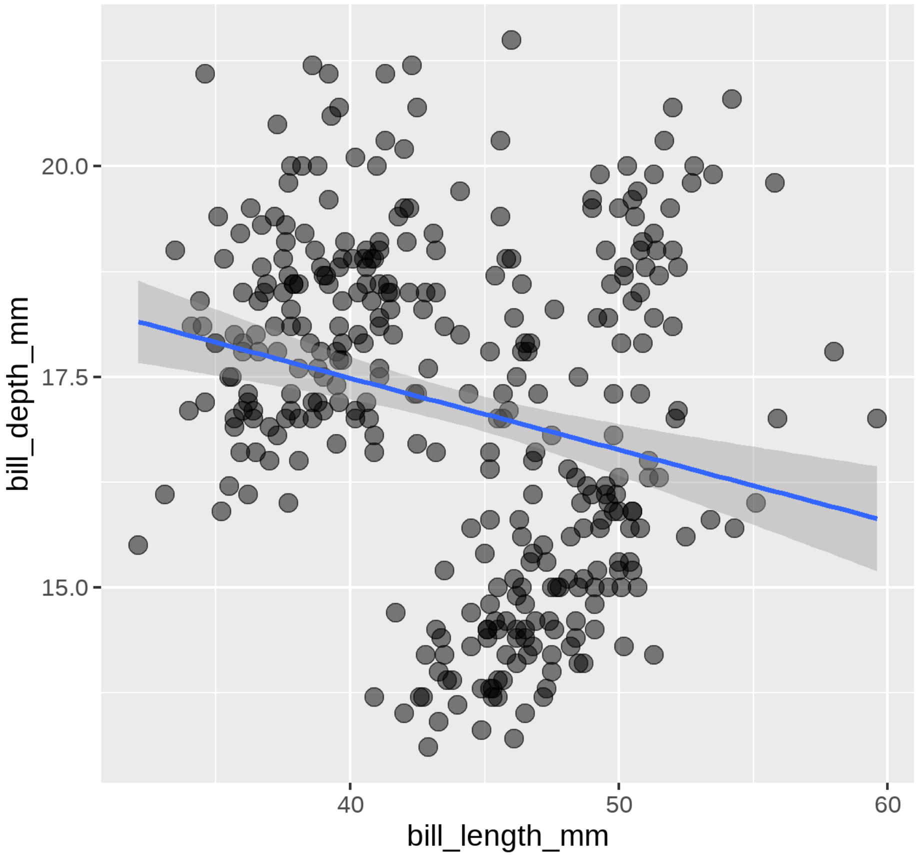
```
head(penguins)
```



```
## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ b
##   <fct>   <fct>           <dbl>         <dbl>            <int> <int> <fct>
## 1 Adelie  Torge~           39.1          18.7              181  3750 male
## 2 Adelie  Torge~           39.5          17.4              186  3800 fema~
## 3 Adelie  Torge~           40.3          18                195  3250 fema~
## 4 Adelie  Torge~           NA            NA                 NA    NA <NA>
## 5 Adelie  Torge~           36.7          19.3              193  3450 fema~
## 6 Adelie  Torge~           39.3          20.6              190  3650 male
## # ... with 1 more variable: year <int>
```

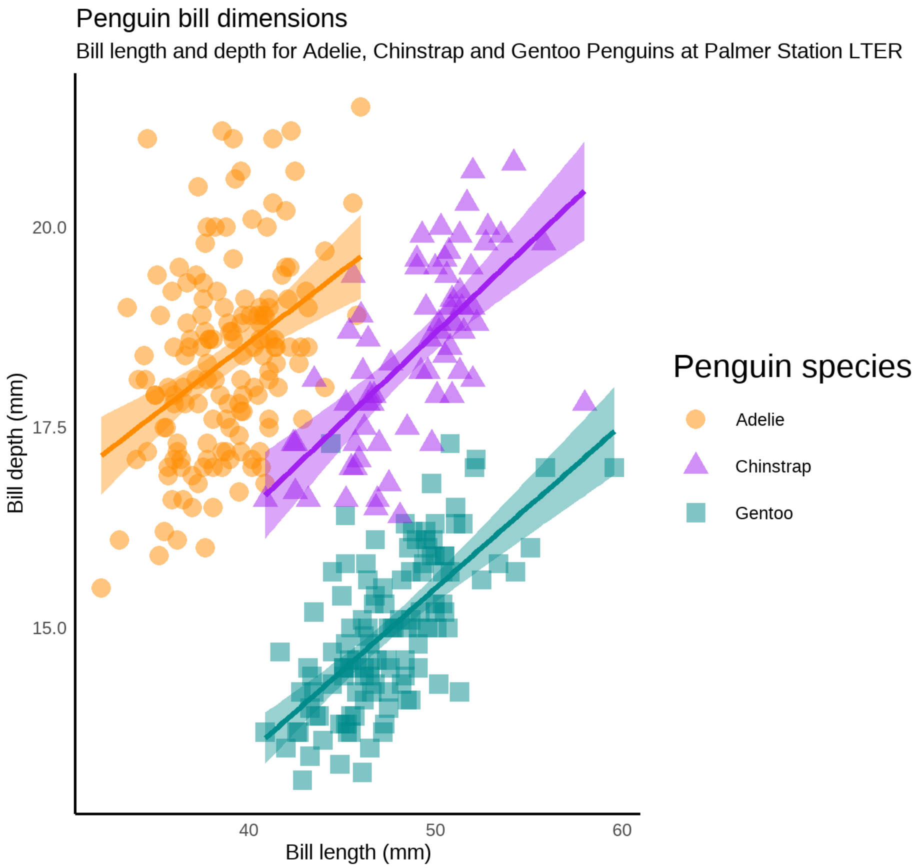
Example of Insights

```
ggplot(data = penguins,
    aes(x = bill_length_mm,
       y = bill_depth_mm)) +
  geom_point(
         size = 3,
         alpha = 0.5) +
  geom_smooth(method = "lm", se = T
```



Example of Insights

```
ggplot(data = penguins,
                    aes(x = bill_length_mm,
                       y = bill_depth_mm,
                       group = species)) +
  geom_point(aes(color = species,
             shape = species),
         size = 3,
         alpha = 0.5) +
  geom_smooth(method = "lm", se = T
  scale_color_manual(values = colors)
  scale_fill_manual(values = colors)+
  labs(title = "Penguin bill dimensions",
     subtitle = "Bill length and depth for
     x = "Bill length (mm)",
     y = "Bill depth (mm)",
     color = "Penguin species",
     shape = "Penguin species") +
  theme_custom()
```



Workshop

- Getting to know R
- Weekly workshops are your **best** way to learn
- Short quizzes to test your understanding

Next Time

A Data Insights walkthrough

Thank you!

Questions?