

Data visualisation in R using crime data

Samuel Langton

5 February 2019

Welcome



@sh_langton

s.langton@mmu.ac.uk

Welcome



@sh_langton

s.langton@mmu.ac.uk

All materials for today are available online.

Web link: https://rpubs.com/langton_

Material: https://github.com/langtonhugh/data_viz_R_workshop

Contents

11.15-11.45

- Intro to data viz
- ggplot2

11.45-12.15

- Live demo

12.15-13.00

- Exercise
- Own data

13.00-14.00

- Lunch

Contents

14.00-15.00

- Intro to spatial viz
- Live demo

15.00-15.15

- Coffee

15.15-16.00

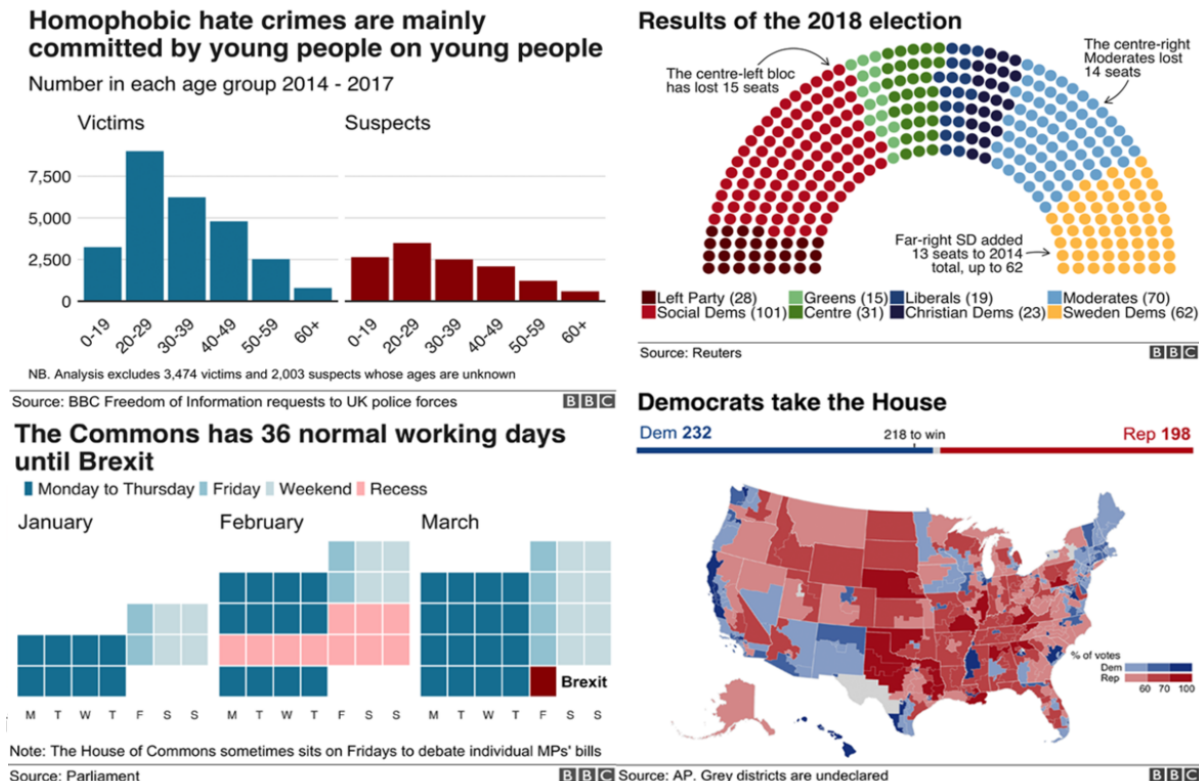
- Exercise
- Own data

Data viz

"The visual representation and presentation of data to facilitate understanding" (Kirk, 2019)

Data viz in R

"The visual representation and presentation of data to facilitate understanding" (Kirk, 2019)



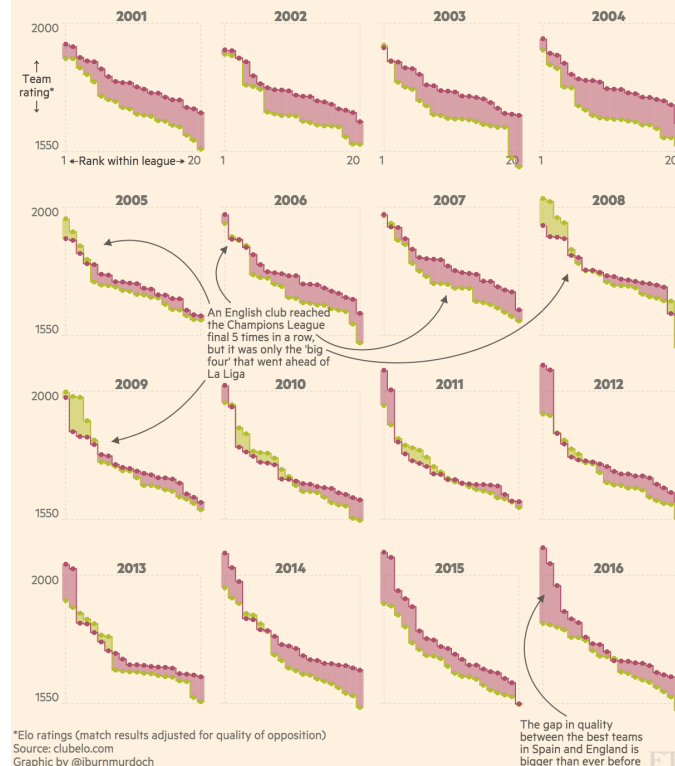
Data viz in R

Was the Premier League ever the strongest league in Europe?

Fans of the English game will point to 2008 & '09, but even then only the very top English sides were stronger than their Spanish counterparts. Today La Liga clubs are stronger from top to bottom

Each string of dots ... represents the clubs in a given country's top league

English team at a given rank was stronger Spanish team at a given rank was stronger

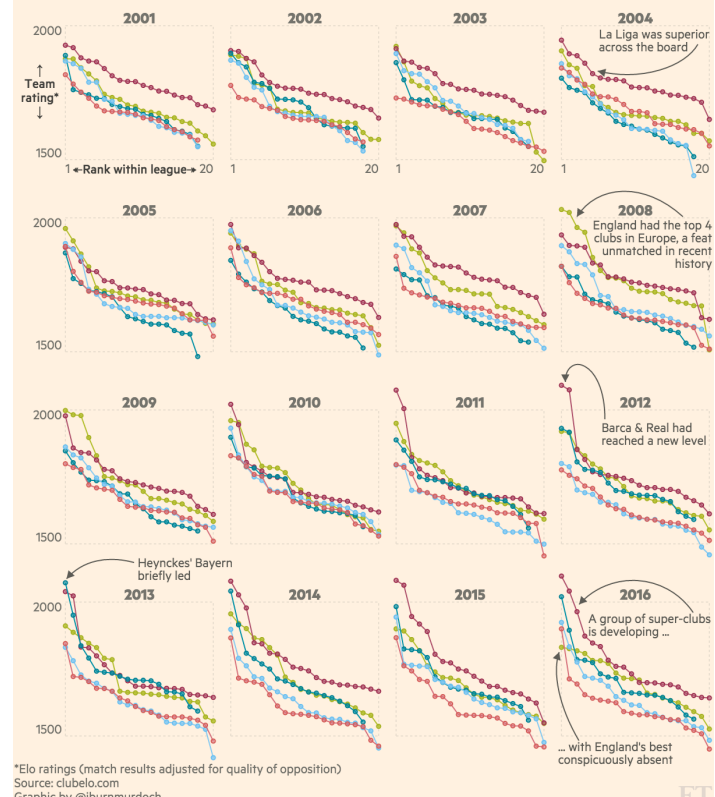


The changing tides of European footballing power

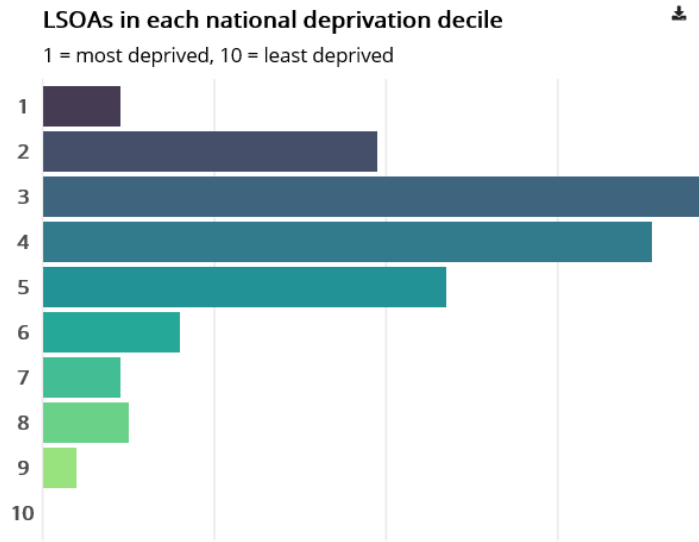
For all the talk of the Premier League being "The Best League in the World™", it's 8 years since its clubs last dominated. Spain's elite & Bayern now lead an elite group, with the English notably absent

Each string of dots ... represents the clubs in a given country's top league

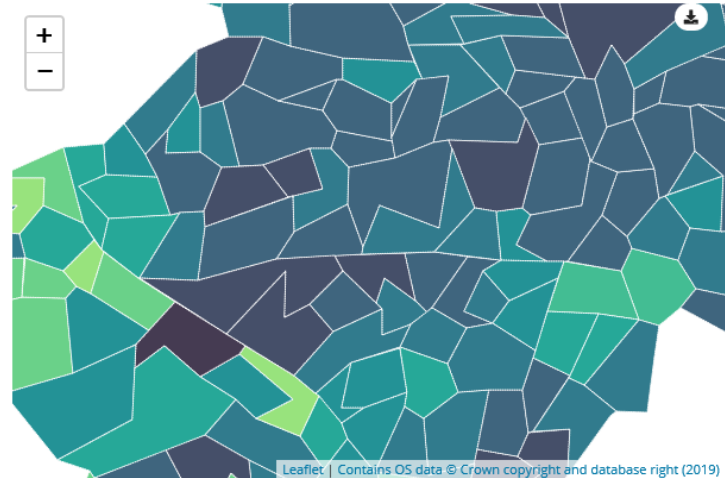
● England ● Spain ● Germany ● Italy ● France



Data viz in R

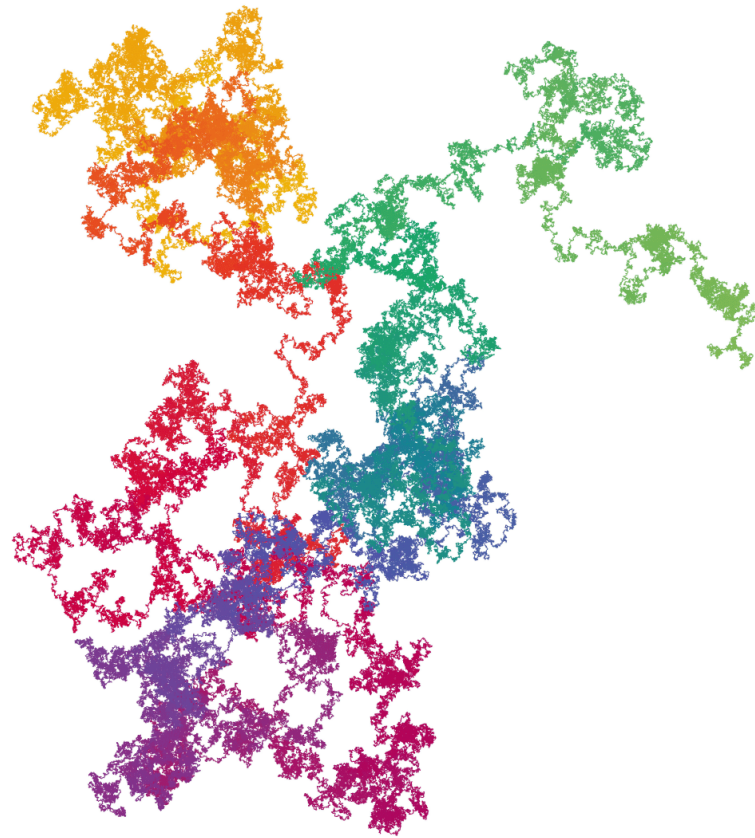


Indices of Deprivation, 2019



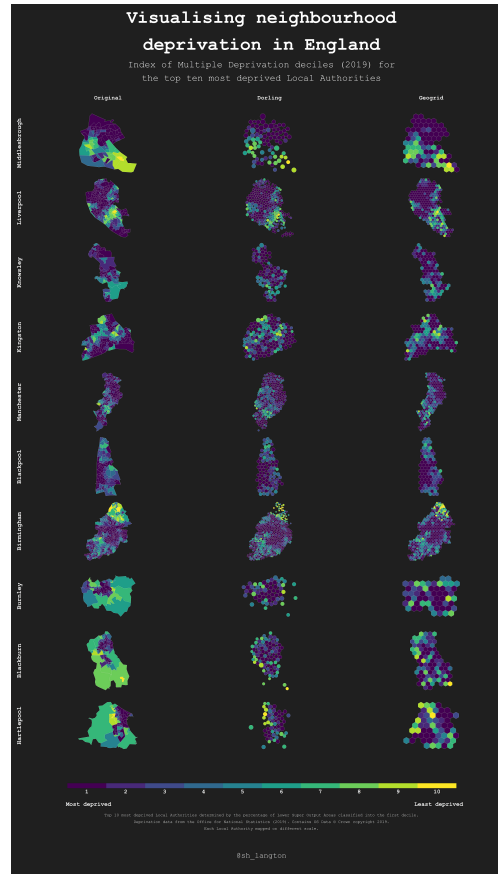
Source: [Trafford Data Lab](#)

Data viz in R



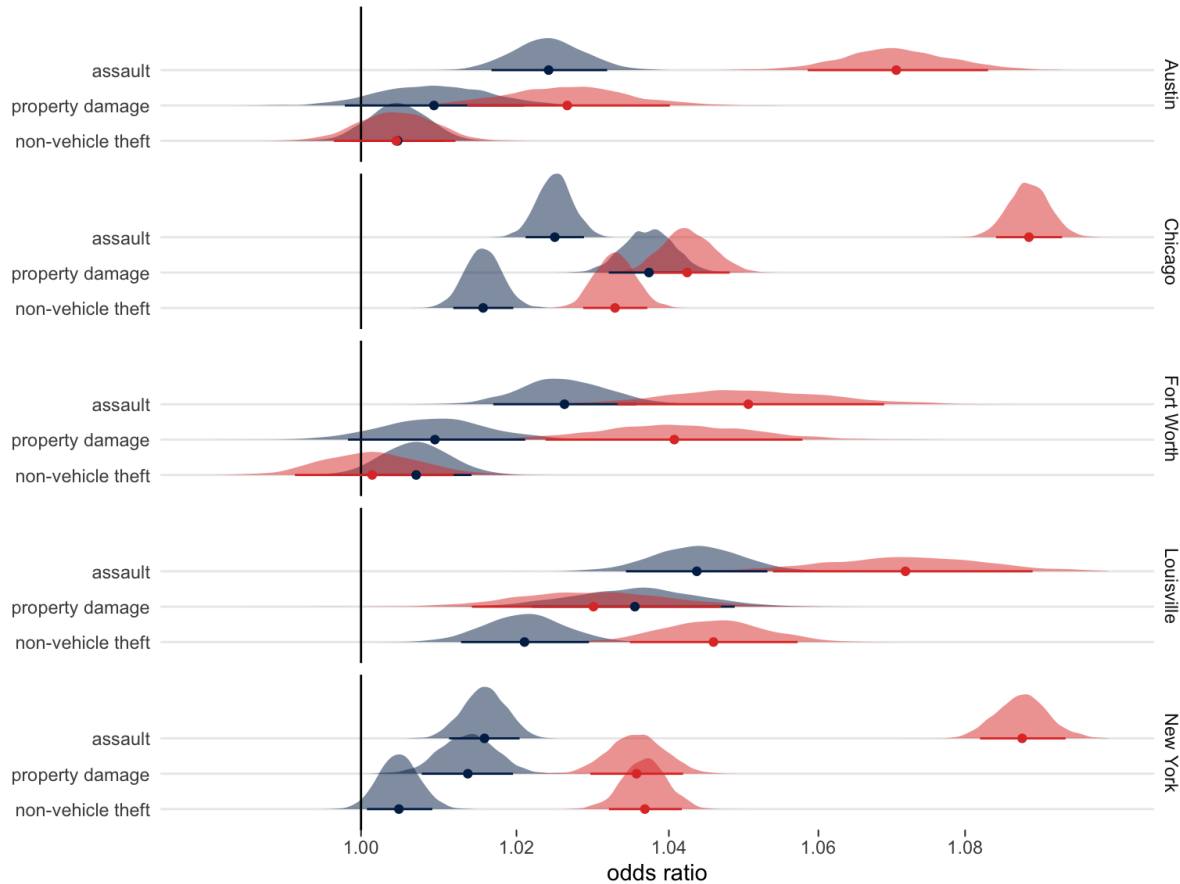
Source: [Nadleh Bremer](#)

Data viz in R



Source: Sam Langton

Data viz in R using crime data



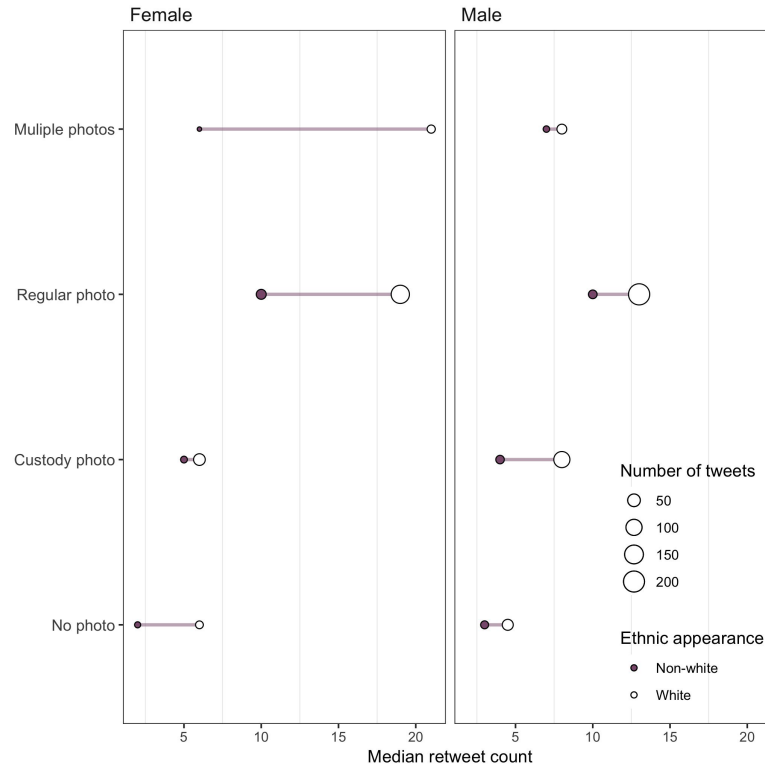
Source: [Matt Ashby](#)

Data viz in R using crime data

Do different images have more retweets?

Tweets with no photo or with a custody image as photo have fewer median retweets than those with a regular (non-custody) photo. Having multiple photos does not seem to get more retweets.

In all cases, missing persons who are white have higher median retweets.



Data: 1008 Twitter appeals for missing persons by Greater Manchester Police Twitter accounts
contact: @r_solymosi

Data viz in R using crime data

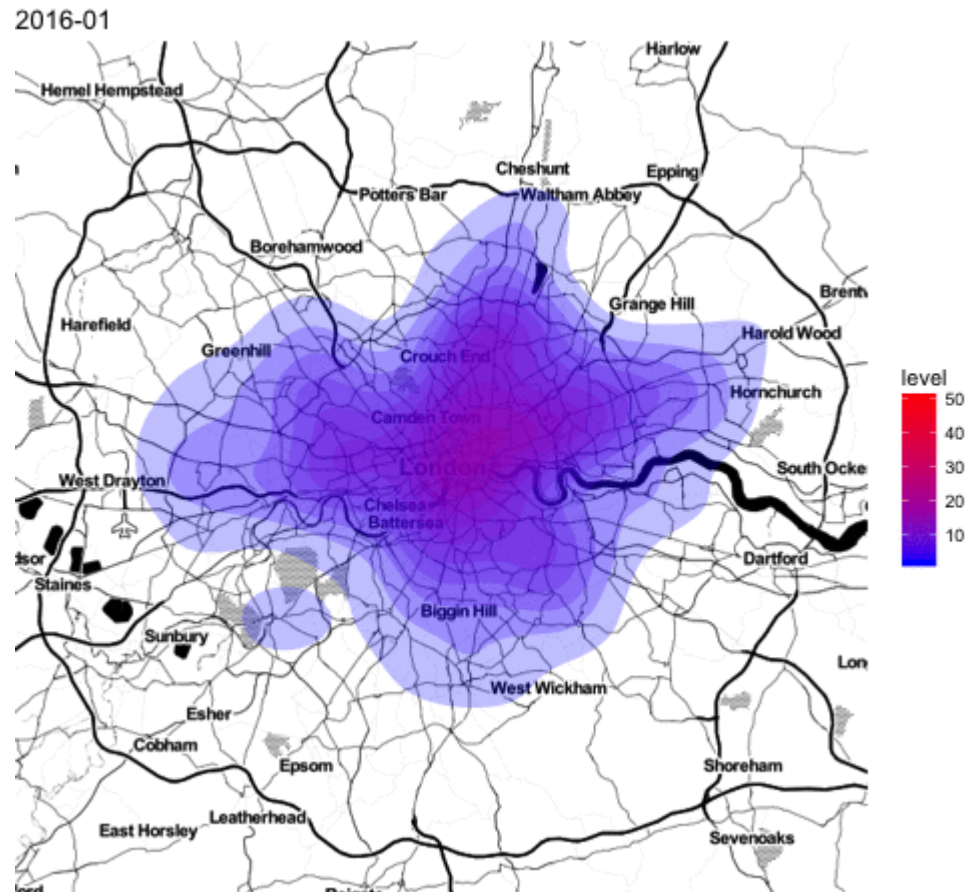
SF has largest concentration of crime near Downtown & Tenderloin

There are also moderate pockets of crime in SOMA & the Mission

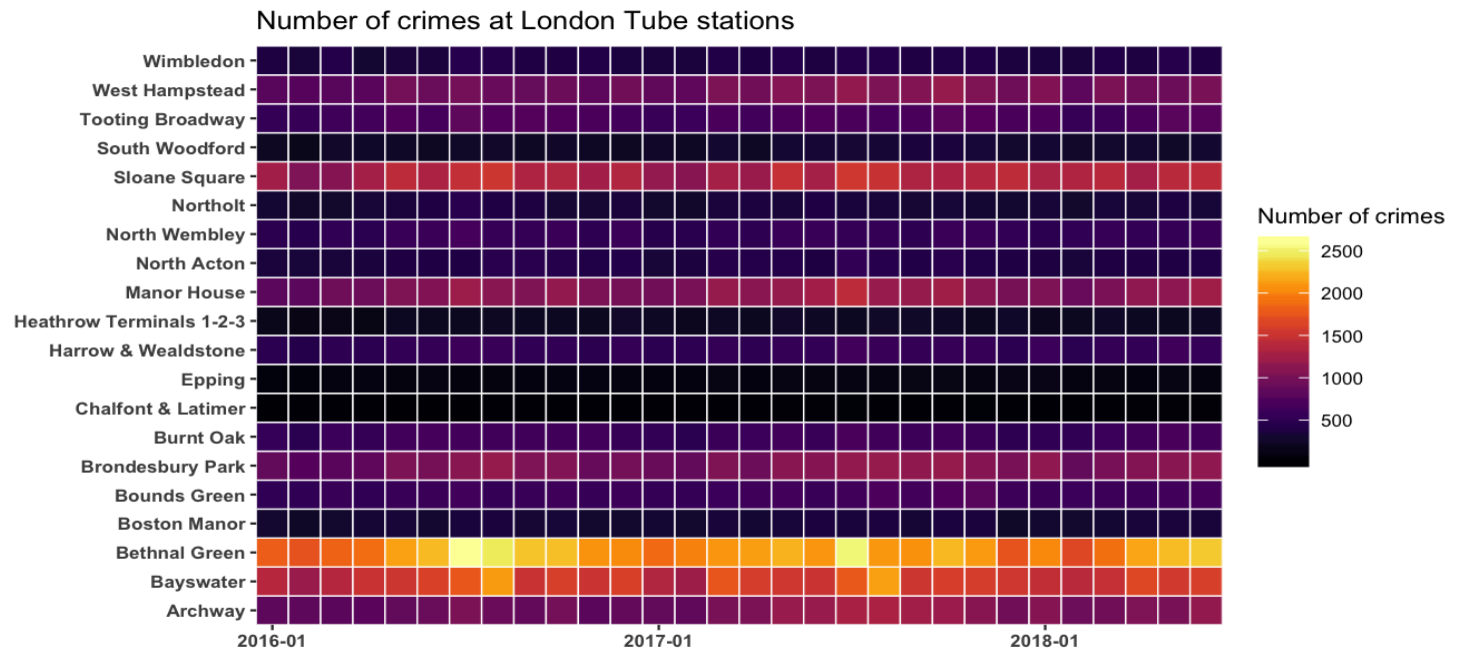


Source: Sharp Insight

Data viz in R using crime data

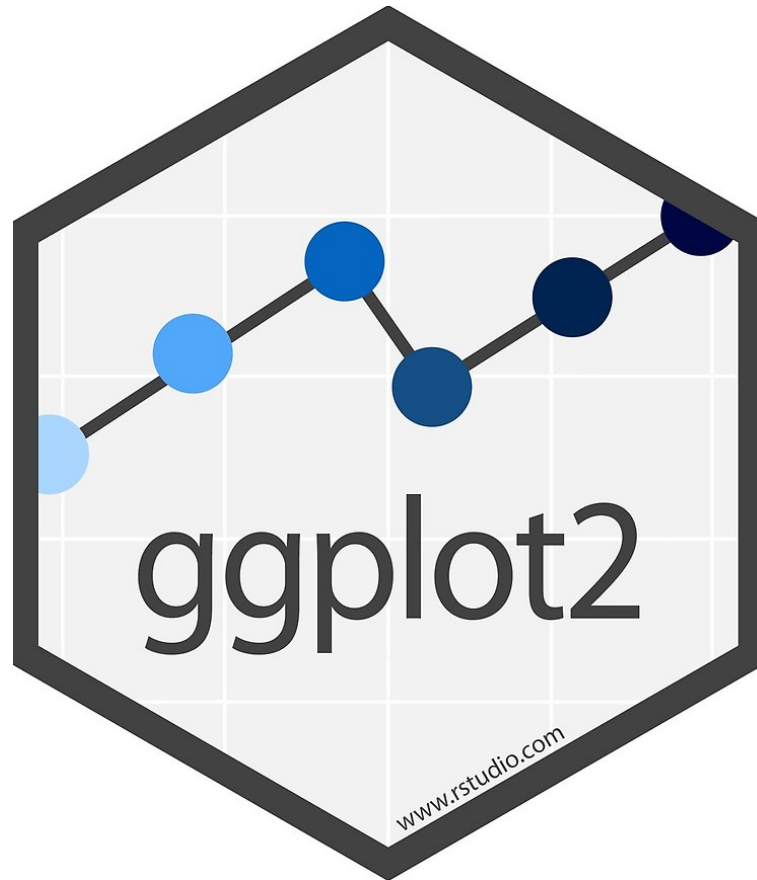


Data viz in R using crime data



Common thread?

Common thread?



ggplot2

- ggplot2 is a package for creating graphics in R based on the **grammar of graphics**.
- A fundamental component of this is that graphics are made up **layers**.
- This way of thinking is reflected in how we write ggplot2 code.



Image source: [Skill Gaze](#)

ggplot2

df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

ggplot2

df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

What is the relationship between var1 and var2?

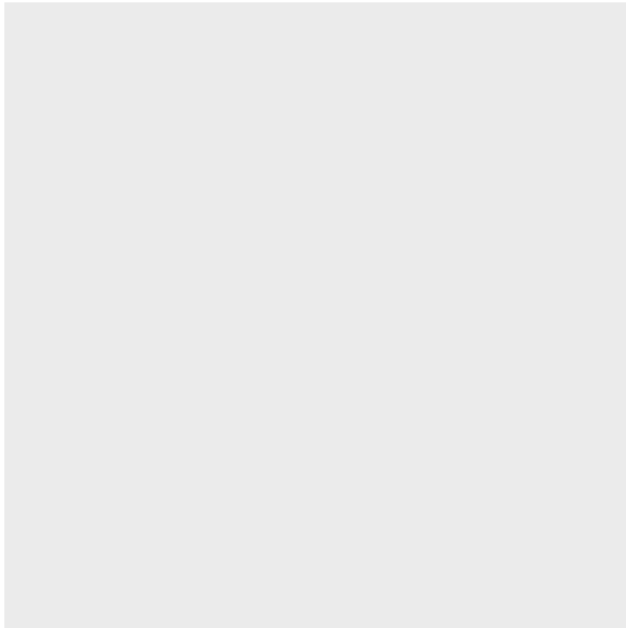
ggplot2: data



Image source: [Skill Gaze](#)

ggplot2: data

```
ggplot(data = df1)
```



df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

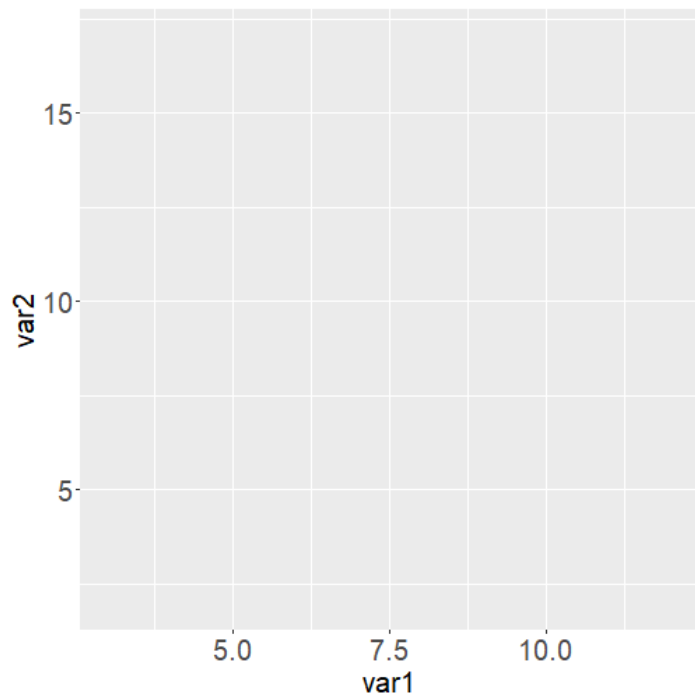
ggplot2: aesthetics



Image source: [Skill Gaze](#)

ggplot2: aesthetics

```
ggplot(data = df1, mapping = aes(x = var1, y = var2))
```



df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

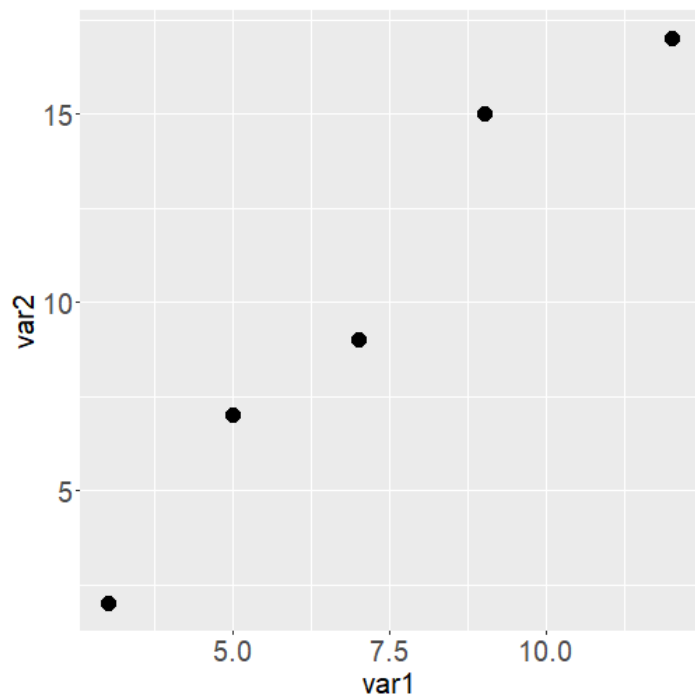
ggplot2: geometries



Image source: [Skill Gaze](#)

ggplot2: geometries

```
ggplot(data = df1, mapping = aes(x = var1, y = var2)) +  
  geom_point()
```



df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

ggplot2: different aesthetics

df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

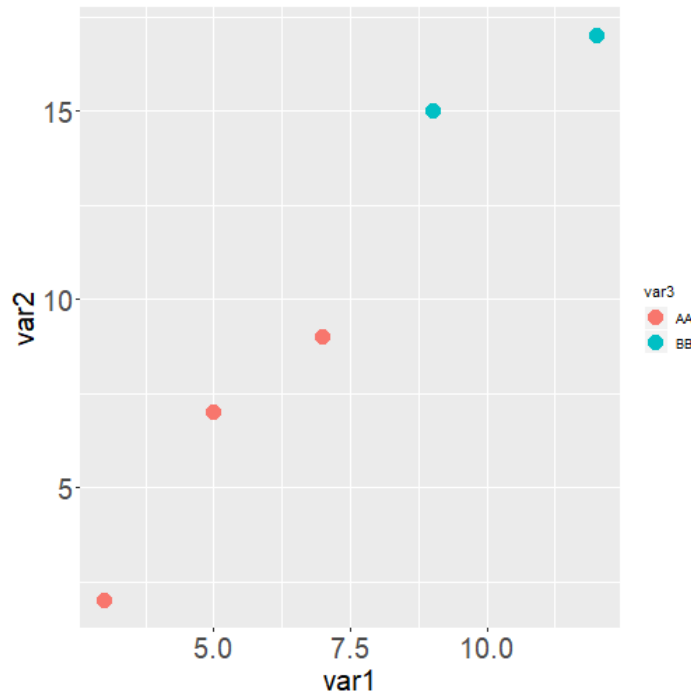
How does var3 factor into this relationship?

ggplot2: different aesthetics

- x
- y
- colour
- shape
- size
- alpha
- linetype
- ...

ggplot2: different aesthetics

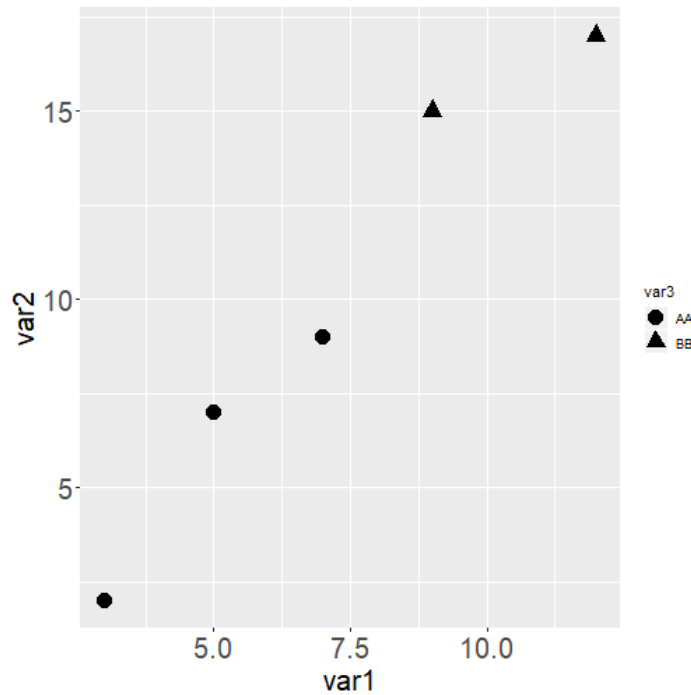
```
ggplot(data = df1, mapping = aes(x = var1, y = var2, colour = var3)) +  
  geom_point()
```



df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

ggplot2: different aesthetics

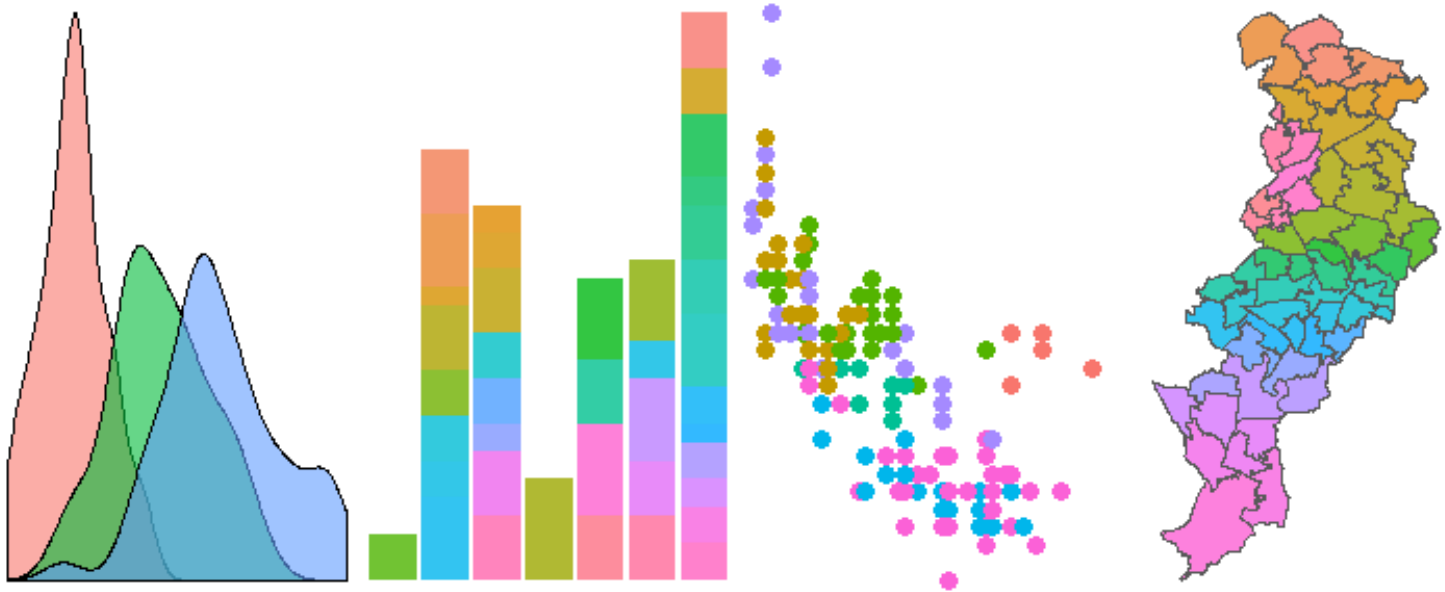
```
ggplot(data = df1, mapping = aes(x = var1, y = var2, shape = var3)) +  
  geom_point()
```



df1		
var1	var2	var3
5	7	AA
3	2	AA
7	9	AA
9	15	BB
12	17	BB

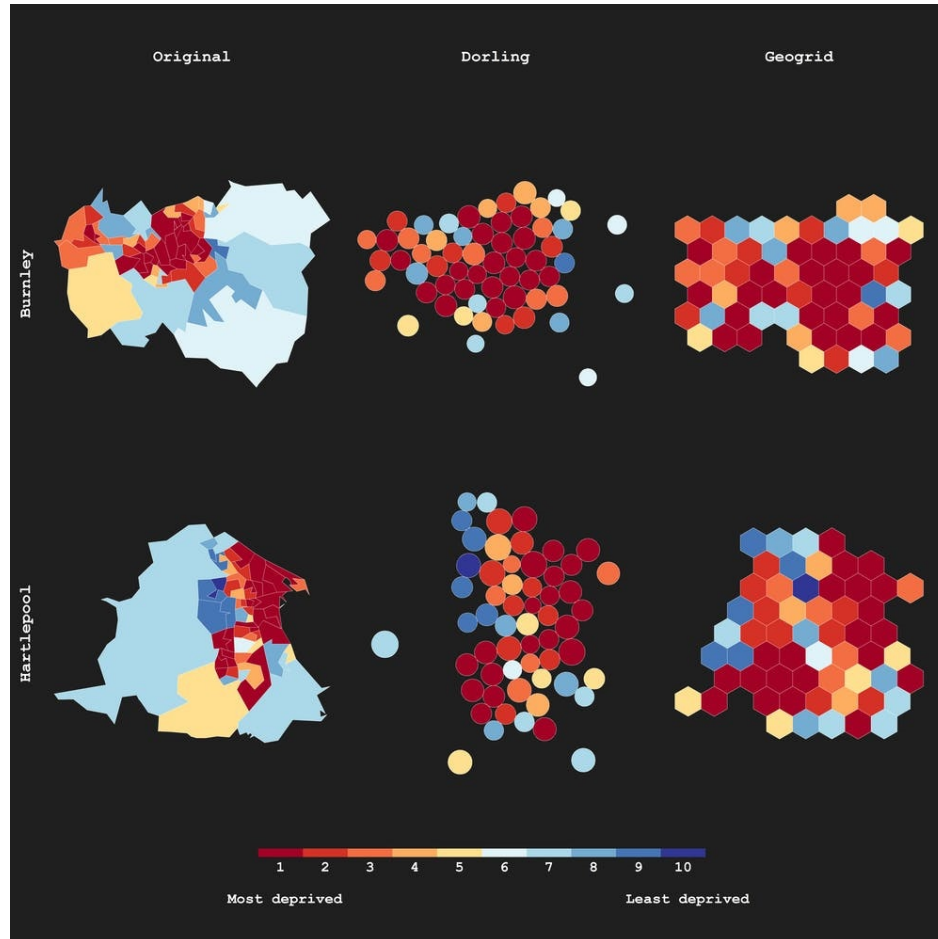
ggplot2: different geometries

- `geom_point()`
- `geom_bar()`
- `geom_density()`
- `geom_smooth()`
- ...



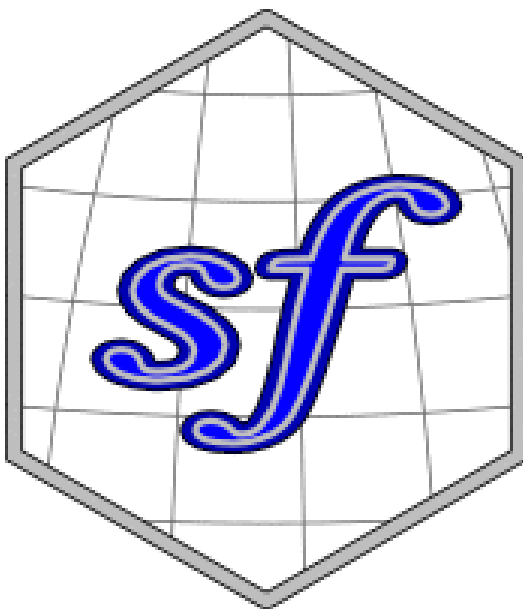
Crime demo

Maps in R



The good news...

- Making a map in R is *very similar* to what we have covered already.
- Using `ggplot2` to create beautiful maps is fairly straightforward, making use of an additional package `sf`.



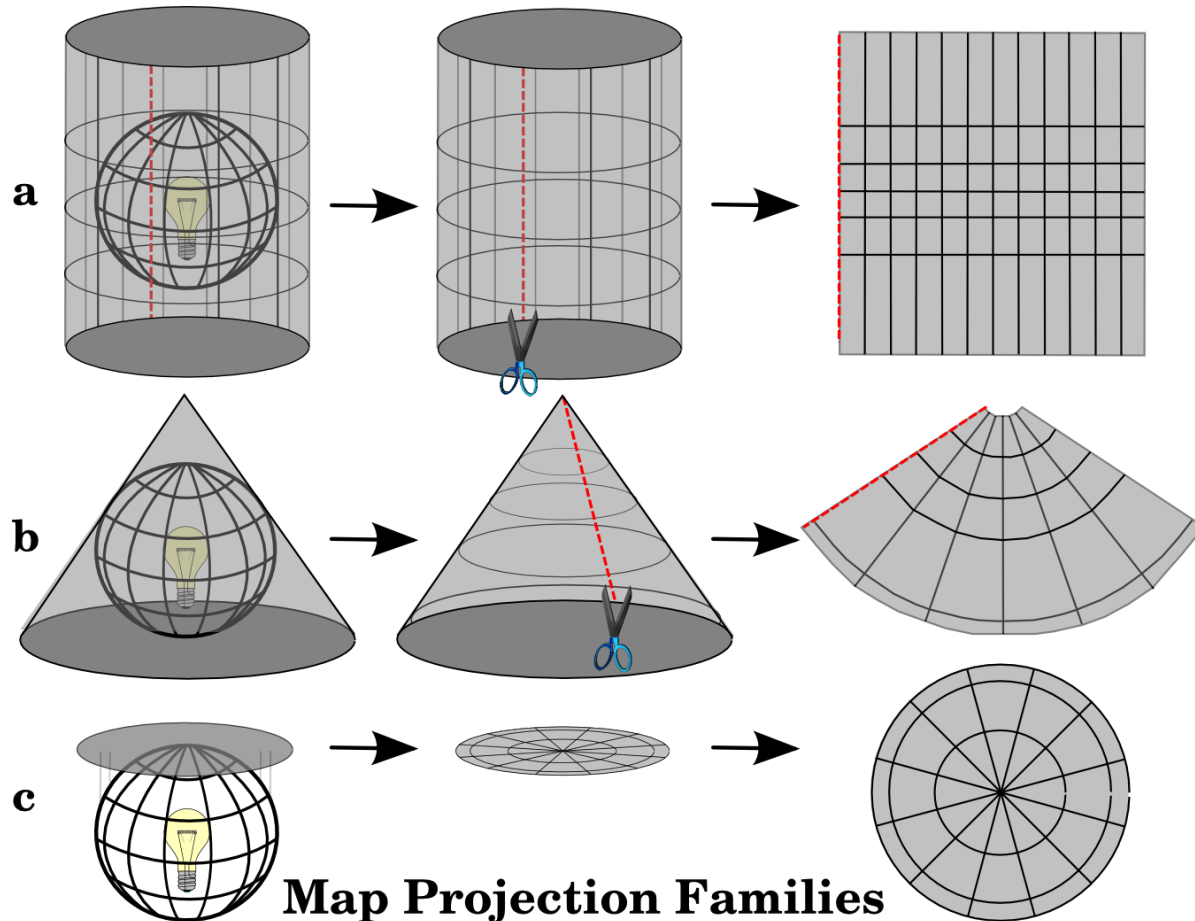
The bad news...

- You might be entering uncharted territory!
- Spatial visualisation necessitates some additional level of understanding in geography and Geographic Information Systems (GIS), including:
 - Spatial data
 - Projection
 - Visualisation issues

The bad news: spatial data

- Today, we will use shapefiles (.shp).
- Shapefiles are a popular format to store geospatial vector data.
- Unlike standard 2D data frames, such as .csv files, shapefiles contains multiple components.
 - .shp
 - .shx
 - .dbf
 - .prj
- But there are other formats, such as *.geojson* or *kml*, which you might come across.
- There is some **debate** over the most appropriate, with each having their own advantages and disadvantages.

The bad news: projections

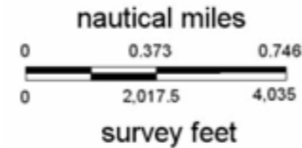


The bad news: visualisation issues

Ten tips to make your maps more 'interesting'

Tip 1: Do not include a scale bar

This will make it much more interesting as your map readers have to guess the distance between objects. One of the main purposes of mapping crime is to compare areas and examine the proximity of objects, so why make it easy for the uninitiated to understand your map? Without a scale bar nobody will have a clue how far things are apart and this gives you the opportunity to have impromptu quizzes or make things up as you are presenting. If you accidentally include a scale bar use a scale that goes; "0 ----- 6.75 ----- 13.25 kilometers" instead of the usual "0 ---5 ---10" or similar. Big complex numbers really impress audiences, and you want to look your most clever.



Source: Jerry Ratcliffe

Crime demo