

# Visualisation: Modelling the World

Richie Morrisroe

July 21, 2019

# Structure

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- This talk is an approach to visualisation
- Not many absolutes
- assumptions of vision
- Assumptions of Statistical Graphics
- Understanding data with Visualisation
- Communicating to others with Visualisation

# What is Visualisation?

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- a tool for understanding the world
- a way to communicate a particular perspective on data
- an adjunct to thought

# Why Visualisation?

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- The eye is really really good at finding patterns in pictures
- in fact, it's so good that it can find patterns that aren't even

there



Figure: What do you see?

# The importance of perspective

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- You can see one of two things in the previous image
- Which of them can depend on what you expect to see
- It can also depend on what your environment contains

# Muller-Lyer

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

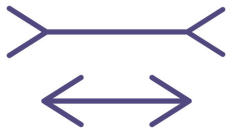


Figure: Which line is longer?

# This illusion doesn't affect everyone similarly

- Europeans and Americans are more susceptible
- Africans are less susceptible
- Possibility that it is due to presence of right angles in urban environments
- appears to be a small difference between urban and rural dwellers
- very very relevant to boxplots (how to lie with boxplots, I guess)

# Who cares?

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- Shows that how we interpret stimuli is not **tabula rasa**
- When you gaze into the image, the image also gazes into you. . .
- We bring our own perception and previous associations into any image <sup>1</sup>

---

<sup>1</sup>anything really, but we're talking about images here.



# When to use Visualisation?

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

Always

# Running Example

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- Property Price Register
  - Kinda a crappy dataset
  - No cleaning or checking done by the authority
  - lots of craziness (1 apartment for 18.6mn)


# Property Price Register

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- We used Google's geocoding service to get more details on each observation
- I updated Shane Lynn's script and ran it on the data up till October 2018
- I also typically break out properties sold for greater than 1e6, as they are often multiple-unit sales (and there's little to no automated way of figuring this out) <sup>2</sup>
- Lots of manual fixing required
- the irish text definitely doesn't help

---

<sup>2</sup>please someone in the audience suggest a better idea 

# Assumptions of Statistical Graphics

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- there are many
- in this section, I'd like to subvert them, in order to make you think

# Line Graphs

Visualisation:  
Modelling  
the World

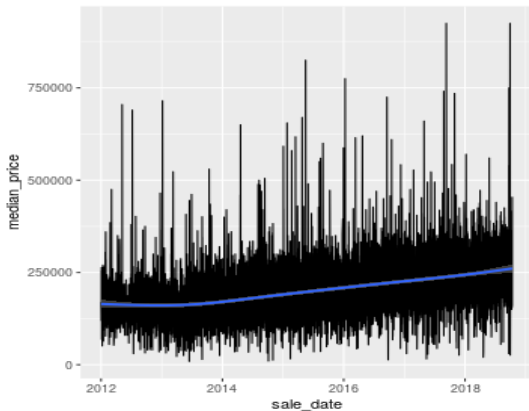
Richie  
Morrisroe

- Normally represent time
- scatterplots don't (always) have the same assumptions
- what is the deepest assumption?

# Median Property Price by Day, Ireland 2011-18

Visualisation:  
Modelling  
the World

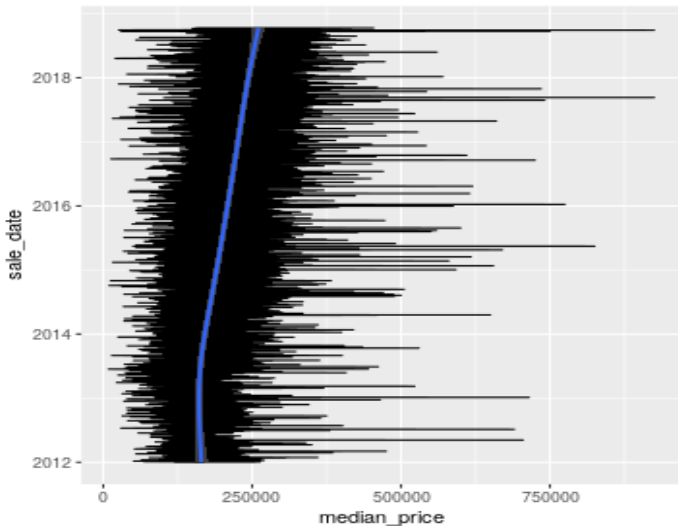
Richie  
Morrisroe



# Flipped Line Chart

Visualisation:  
Modelling  
the World

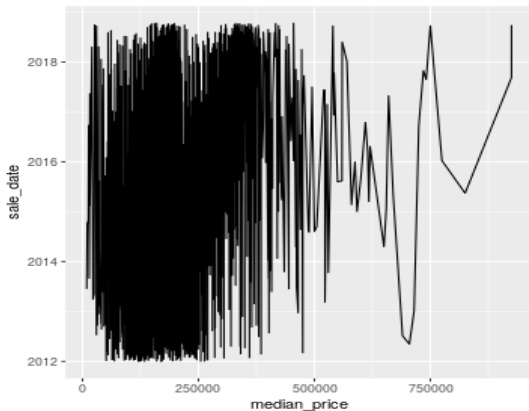
Richie  
Morrisroe



# F-ing Line Chart

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



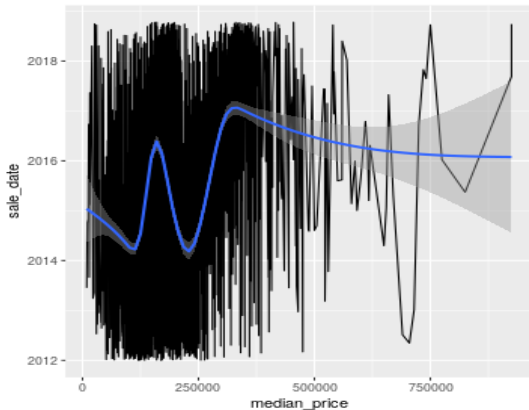
- Here, the violence is that we swap the axes in a fashion only a monster would



# Abusing Standard Assumptions

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



# Scatter plot

Visualisation:  
Modelling  
the World

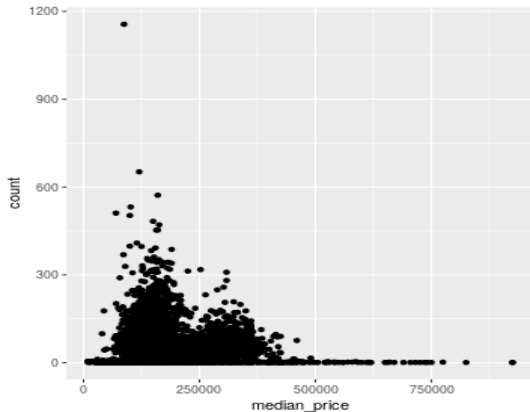
Richie  
Morrisroe

- Also encodes a set of base assumptions
- points nearer to each other in space are more related
- more orientation issues

# Standard Scatter

Visualisation:  
Modelling  
the World

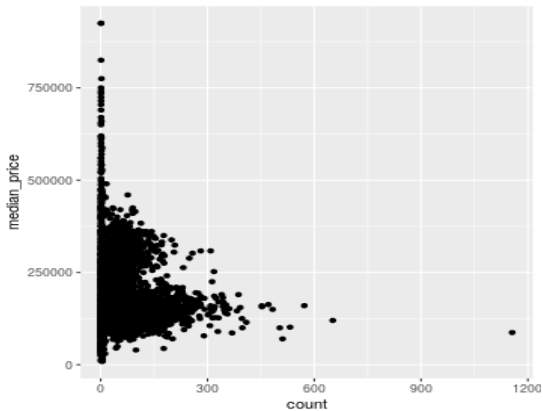
Richie  
Morrisroe



# Flipped Scatter

Visualisation:  
Modelling  
the World

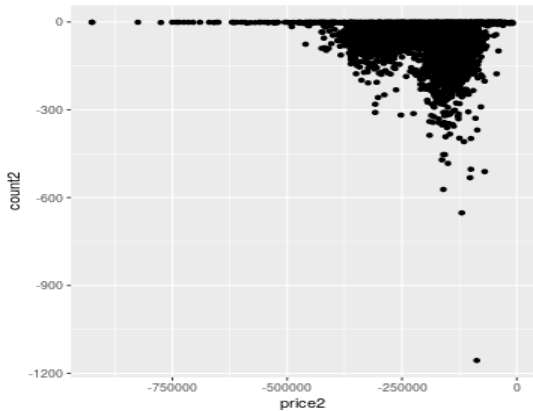
Richie  
Morrisroe



# Other side

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



# What does this tell us?

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- We have a base level of assumptions that we bring to graphics (especially statistical graphics)
- Most of these appear to have been formed by Descartes
- When these assumptions are subverted, expect problems

# Simple Statistical Graphics

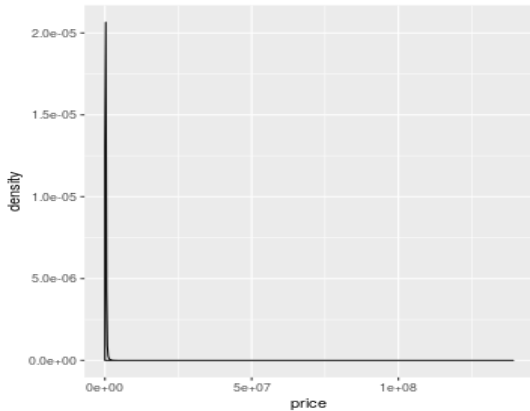
- Graphs excel at showing relations between things
- Consider the difference between quantiles of a variable, and a density plot
- For example, the price of houses:

0%	5079
10%	55000
20%	85000
30%	115000
40%	145000
50%	175000
60%	214000
70%	255505
80%	315000
90%	430000
100%	139165000

# Density Plot

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

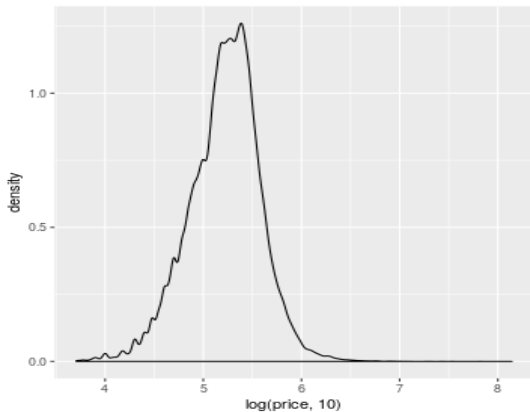




# Better Density Plot

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



# Transformations

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

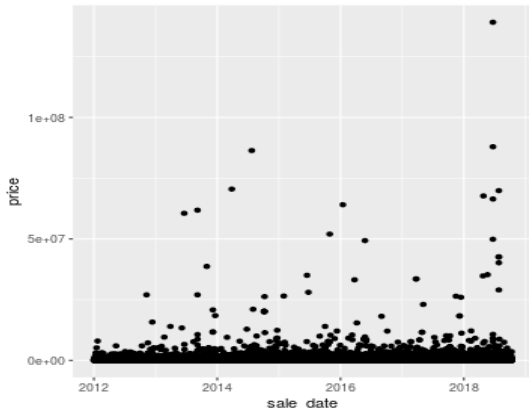
- Useful to get a better sense of the data
- Have a bunch of assumptions (what's the log of -1)
- Can be used to deceive very, very easily
- Really really useful in everyday practice

# Getting the sense of things

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- Picking the right visualisation for the data is important



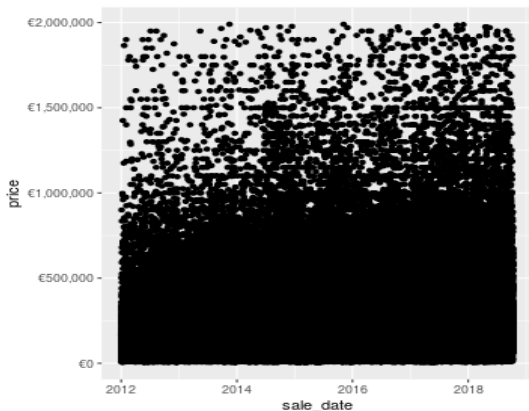
- is this a good plot?
- does this depend on the number of points?

# Cleaning the Data

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

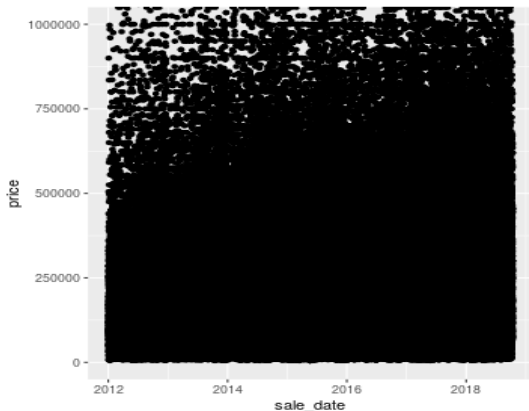
- Let's say we remove all properties with prices greater than 2mn



# More Data Cleaning

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

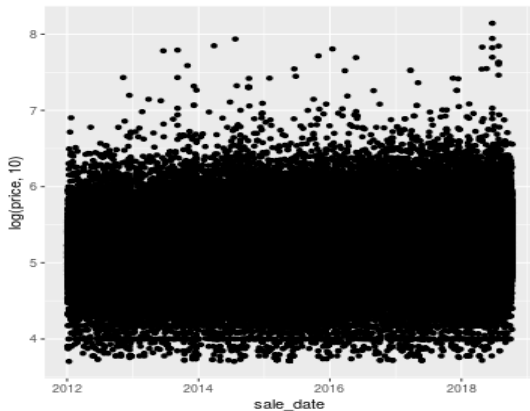


- Better or worse?

# Transformations Help

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



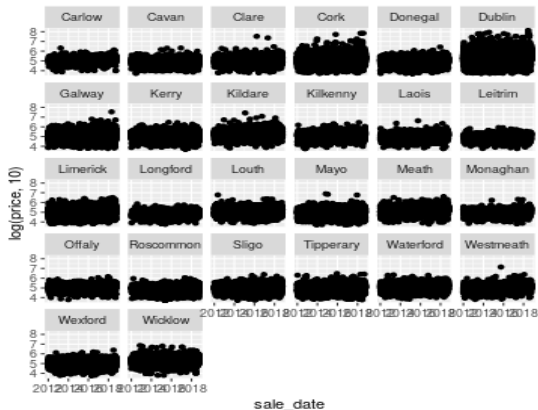
- Note the log 10 base
- Some of you may be able to convert from base 2.718, but I missed that class in school
- Still crap though

# No data is an island

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- The first obvious thing is to split by county, right?



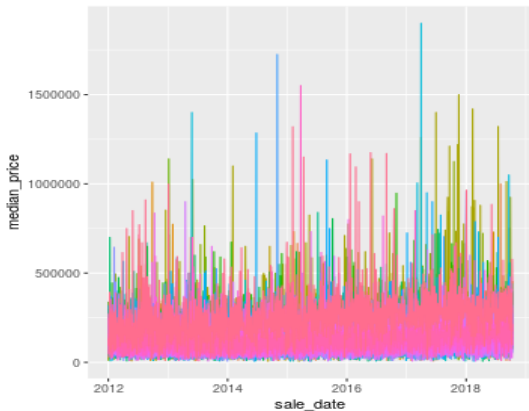
- Oh look, it's lot of little boxes of crap :(

# Summarisation

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- The obvious answer is summarisation

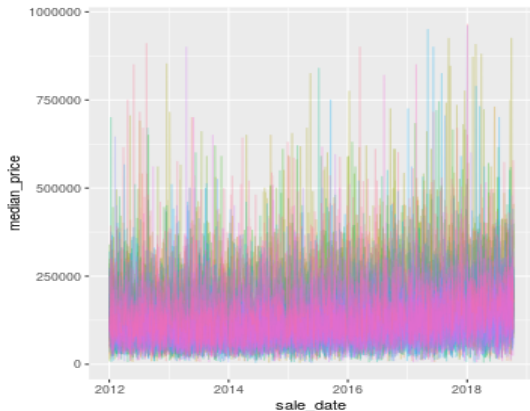




# Reducing Alpha kinda works...

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



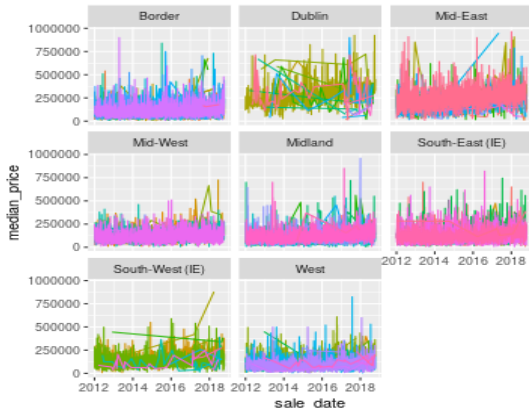
- But really just washes the whole thing out

# A redundant faceting variable

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- We just group by a higher level variable



- Much clearer :)

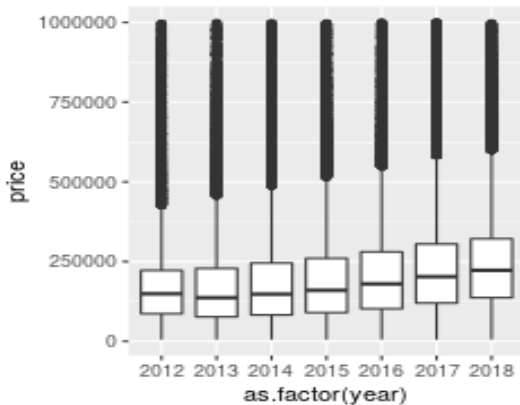
# WTF?

- This is one of the major advantages of visualisation:
  - it helps to (dis)confirm your assumptions
  - given that we have too many lines in the various groupings, we know that something has gone horribly wrong
  - in this case, it's a mismatch between two different types of data

# Distributions (i.e. boxplots)

Visualisation:  
Modelling  
the World

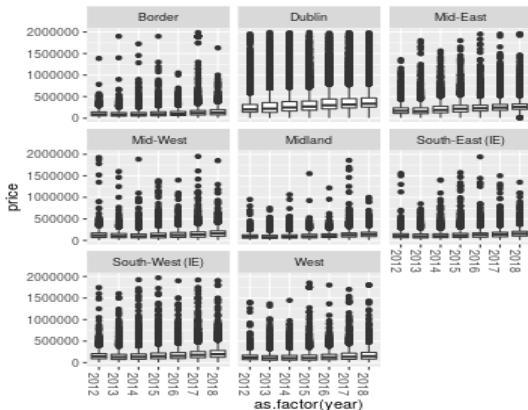
Richie  
Morrisroe



# Faceting, redux

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

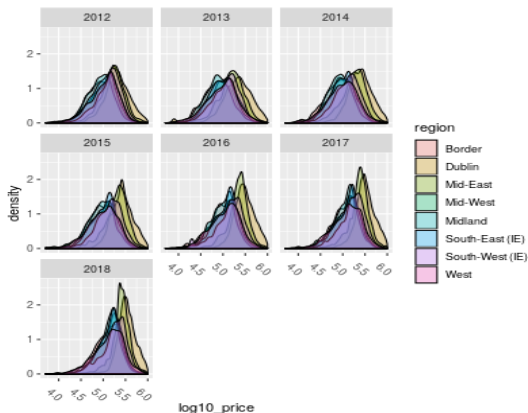


- This actually works (for me, at least)
- can you explain this to a sales-person?

# Distributions over Time, Redux

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



- This is much, much better
- I definitely don't think I'd try to explain it to a business/sales person

# Spatial vs Temporal

Visualisation:  
Modelling  
the World

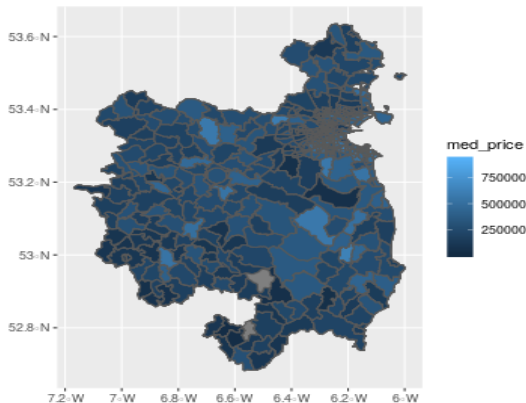
Richie  
Morrisroe

- line plots vs maps
- time versus space
- both provide insight into
- pick one, difficult to do both

# Line plots ignore space, maps ignore time

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



- There's a real problem of scale here, in that Dublin City is both responsible for much of the population, but is invisible

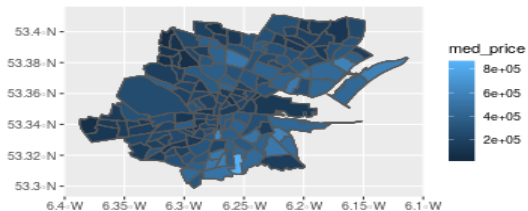




# Dirty Oul Town

Visualisation:  
Modelling  
the World

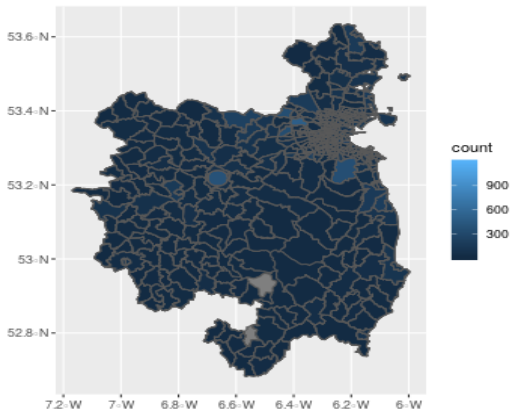
Richie  
Morrisroe



# Counts tell a different story

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

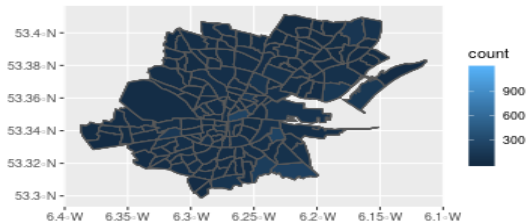


- Outliers make the map useless

# Dublin City (again)

Visualisation:  
Modelling  
the World

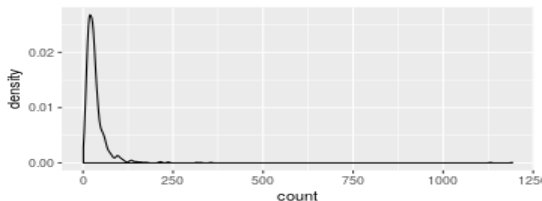
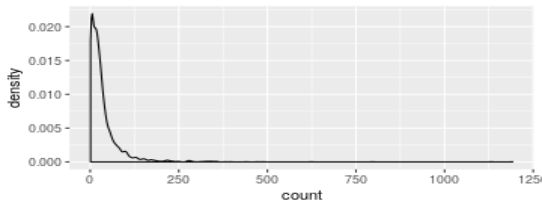
Richie  
Morrisroe



# Density Plots to help maps

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

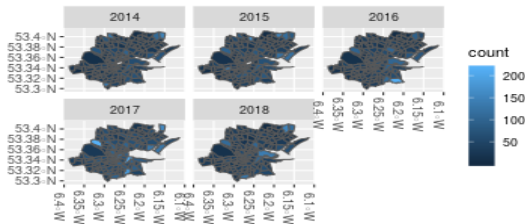


- A tiny proportion of electoral districts drive the uselessness of the maps

# Maps over Time

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

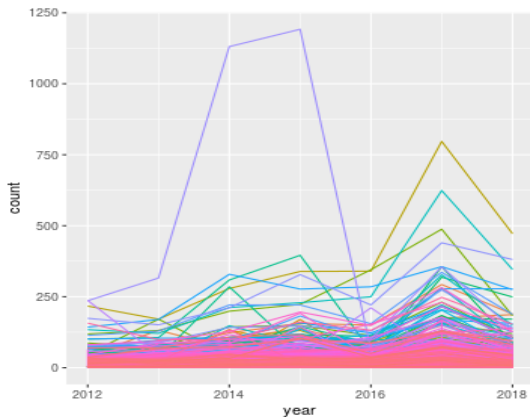


- Just doesn't work
- Even when I account for the outliers, it still doesn't work.

# Lines for Time

Visualisation:  
Modelling  
the World

Richie  
Morrisroe



- This shows the trend plus outliers
- Much more useful
- lose the spatial dimension

# Interactivity and Dashboards

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- Can show both time and space
- for reporting, these are essential
- Much more effort from a software-engineering perspective <sup>3</sup>

---

<sup>3</sup>for me, at least

# Performative vs Presentation

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- Two types of graphs:
  - for yourself
  - for other people (and different audiences need different things)



# Performative Graphics

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- These are used to help you understand a problem
- typically created in an iterative fashion
- often move from data transformation to visualisation and back again (like this talk)

# Presentation Graphs

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- To some extent, your job with presentation visualisations is to tell a story
- hopefully, it will be nuanced, but that isn't a requirement <sup>4</sup>
- Often good to show smooths as opposed to raw data
- raw data is often ugly
- need for care here, as this should only be done where there is a clear effect

---

<sup>4</sup>and in fact, it may be better to remove all nuance from the presentation and provide a longer document with all the failed approaches and hacking needed to actually reproduce your results

# Advice

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- As few as possible
- One clear message
- Repeat yourself
- Remove nuance

# As few as possible

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- There should be no extraneous graphs
- Each graph should have a clear purpose
- Smooths are really effective

# One Clear Message

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- You should only be telling one story at a time
- People are easily confused
- Especially in an oral presentation
- Backup docs should contain nuance

# Repeat Yourself

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- This is the key to helping people retain information
- This is easier once you know the story
- Say what you want to say, say it, then say what you said

# Remove Nuance

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- This varies by audience
- Salespeople may just want the results
- colleagues may want to see the code
- most people just want a high level explanation
- Nuance should be present, just not in a presentation

# Conclusions

- Everyone bring assumptions to visualisations
- Make sure that you take advantage of this
- Visualisation is primarily a tool for communicating with yourself
- Iterative process, even bad graphs can teach you something
- Secondly, it's a tool for communicating with others
- When using visualisations with others, keep it simple



# More Info

Visualisation:  
Modelling  
the World

Richie  
Morrisroe

- My property article here
- My repository for this talk
- My crazy long notes file with most of my analyses
- the data itself

# sessionInfo

Visualisation:  
Modelling  
the World

Richie  
Morrisroe