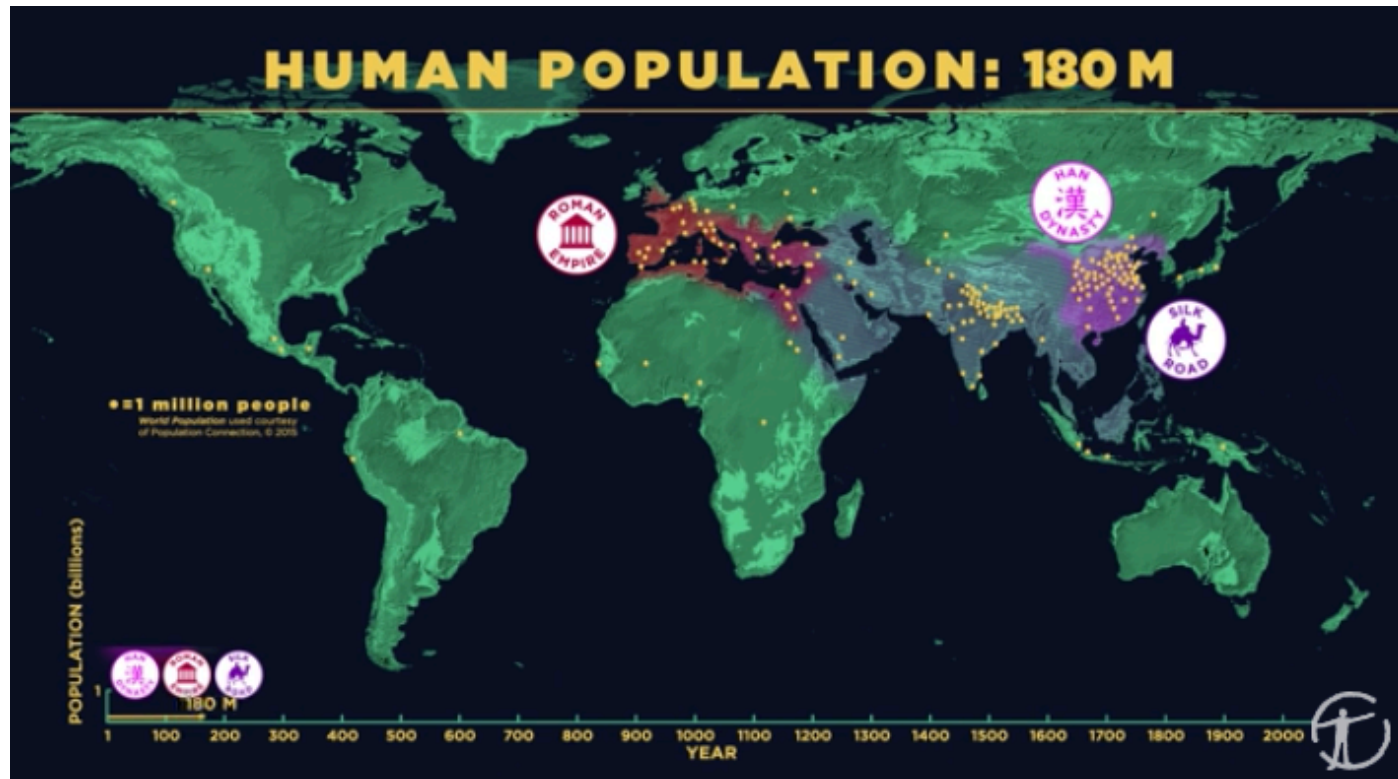




# FIT5147 Data Exploration & Visualisation

Kim Marriott

# Hall of Fame/Shame (Tue)



Girish Bhatta

[https://www.youtube.com/watch?v=PUw mA3Q0\\_OE](https://www.youtube.com/watch?v=PUw mA3Q0_OE)

# Hall of Fame/Shame

SANFORD AND SELNICK

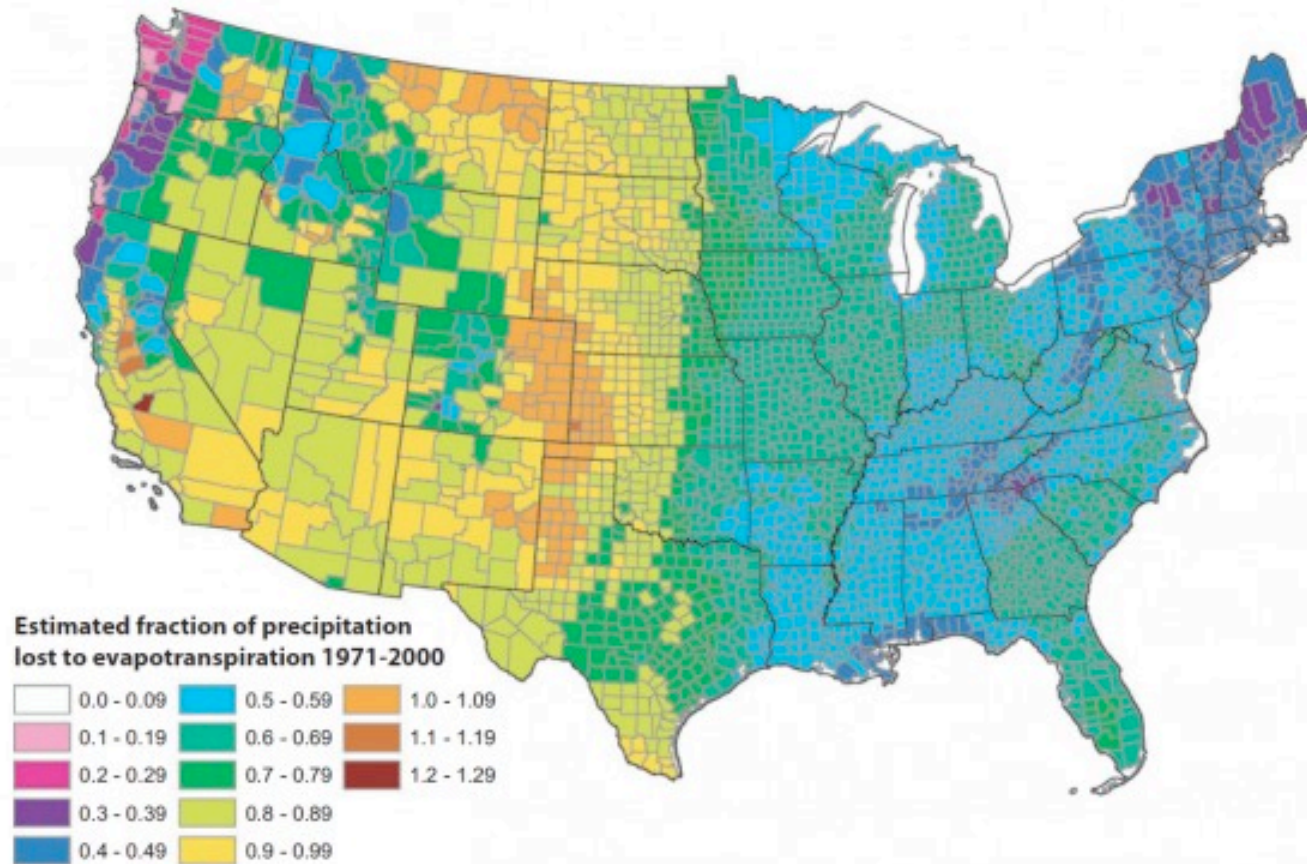


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Yalong Yang

<https://eagereyes.org/basics/rainbow-color-map>

# Syllabus

Week	Lecture material	Lab/Tute
. 1	Visual analytics; Tools for data exploration	Intro to Tableau; R; D3
. 2	<u>Visualisation</u> of tabular data	Advanced graphics with R
. 3	Analysis of trends & patterns in tabular data	Interactive graphics with R
. 4	<del>Temporal data analysis; Tools for exploring data trends</del>	<del>Temporal data analysis; Tools for exploring data trends</del>
. 5	Spatial analytics	<u>MapBox</u> ; Data Exploration Project feedback
. 6	<del>Network data analysis &amp; visualisation</del>	<del>Relational data analysis; Text and text analytics with R</del>
		R
. 7	Textual data analysis & <u>visualisation</u>	Data Exploration Project Feedback
Break		
. 8	<u>Visualisation</u> design methodology	Five design sheet <u>visualisation</u> design methodology
. 9	Human visual system	Introduction to D3
. 10	Visual communication	More D3; Data Vis Project Feedback
. 11	Interactive data <u>visualisation</u>	Data Vis Project Presentations
. 12	History and <u>future</u> of data <u>visualisation</u>	Five design sheet <u>visualisation</u> design methodology

# Analytics for Spatial Data

Once the amount of data gets very large we can't visualise all of it

We also want help in drawing conclusions from the data

**Analytics** helps.

What kinds of analytics are there?

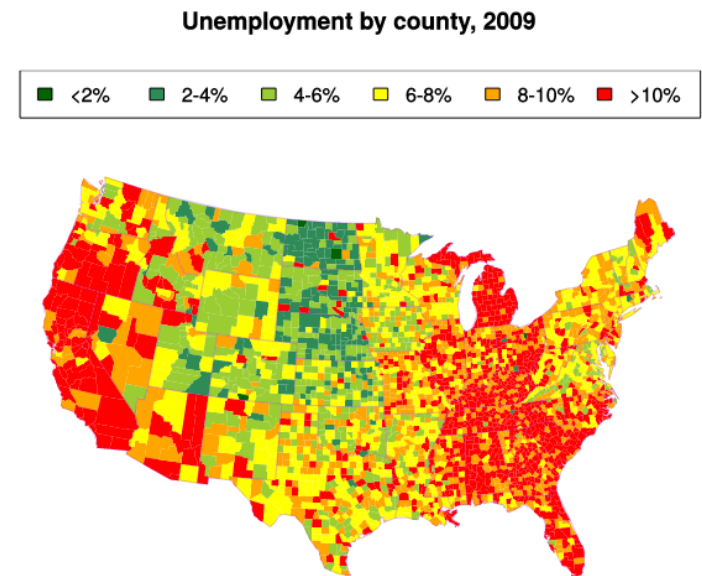
# Choice of Classes

Need to choose classes in classed choropleth maps and range graded proportional symbol maps

Typically 4-5 classes but in exploration may use more

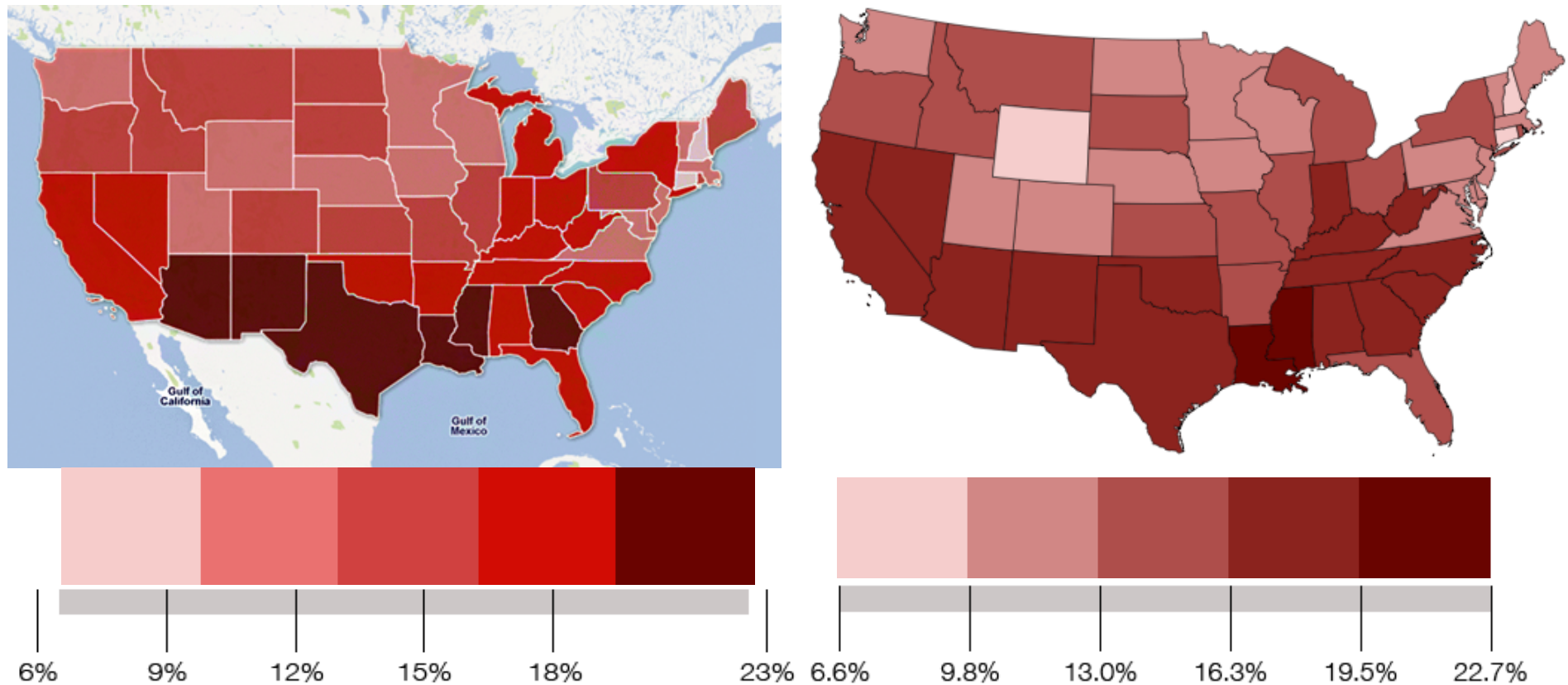
Choosing class intervals

- Equal (Size) Interval
- Equal Frequency
- Jenks Optimisation





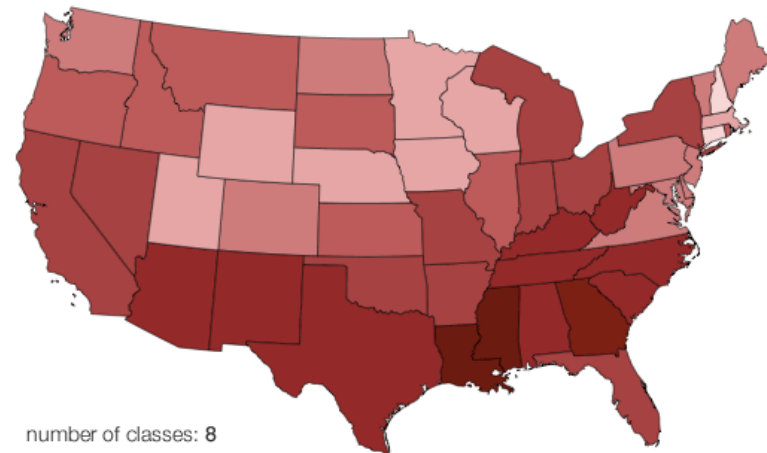
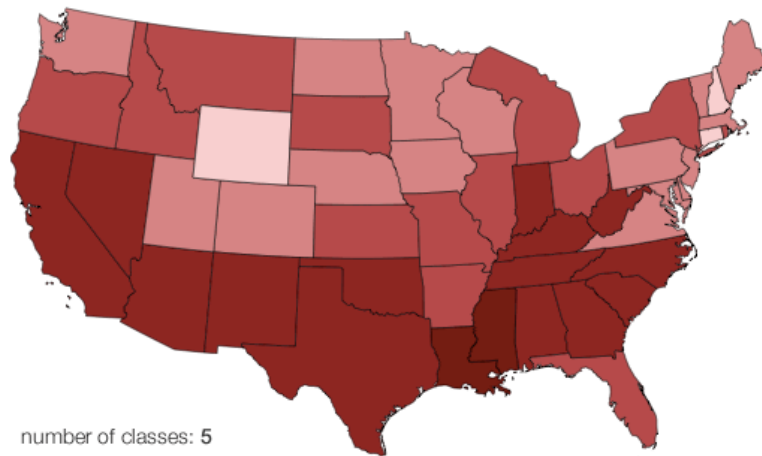
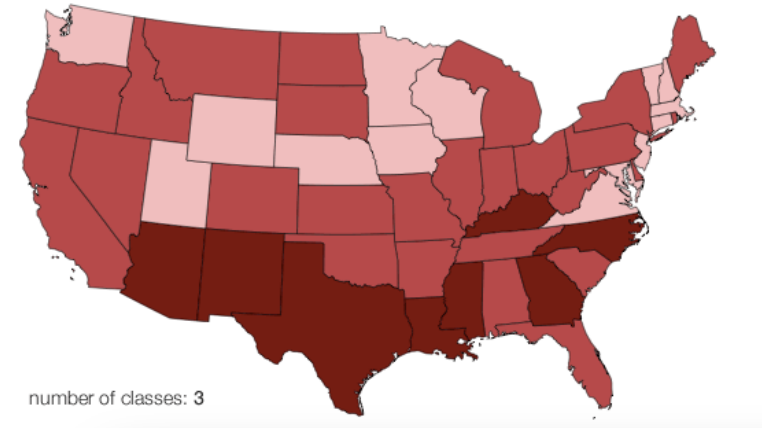
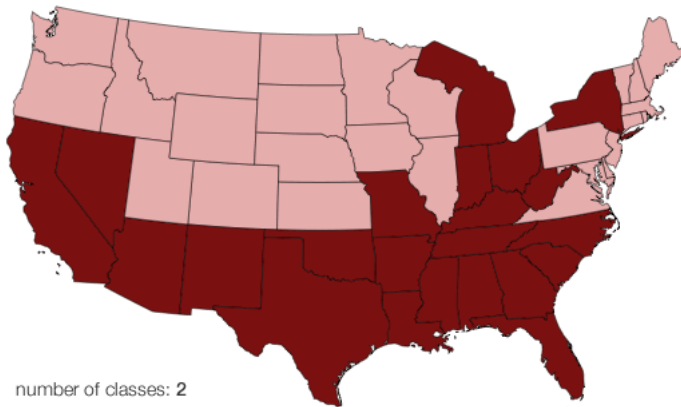
# Example: US Poverty



Guardian data blog – number of people below poverty line in US states

Taken from <https://vis4.net/blog/posts/choropleth-maps/>

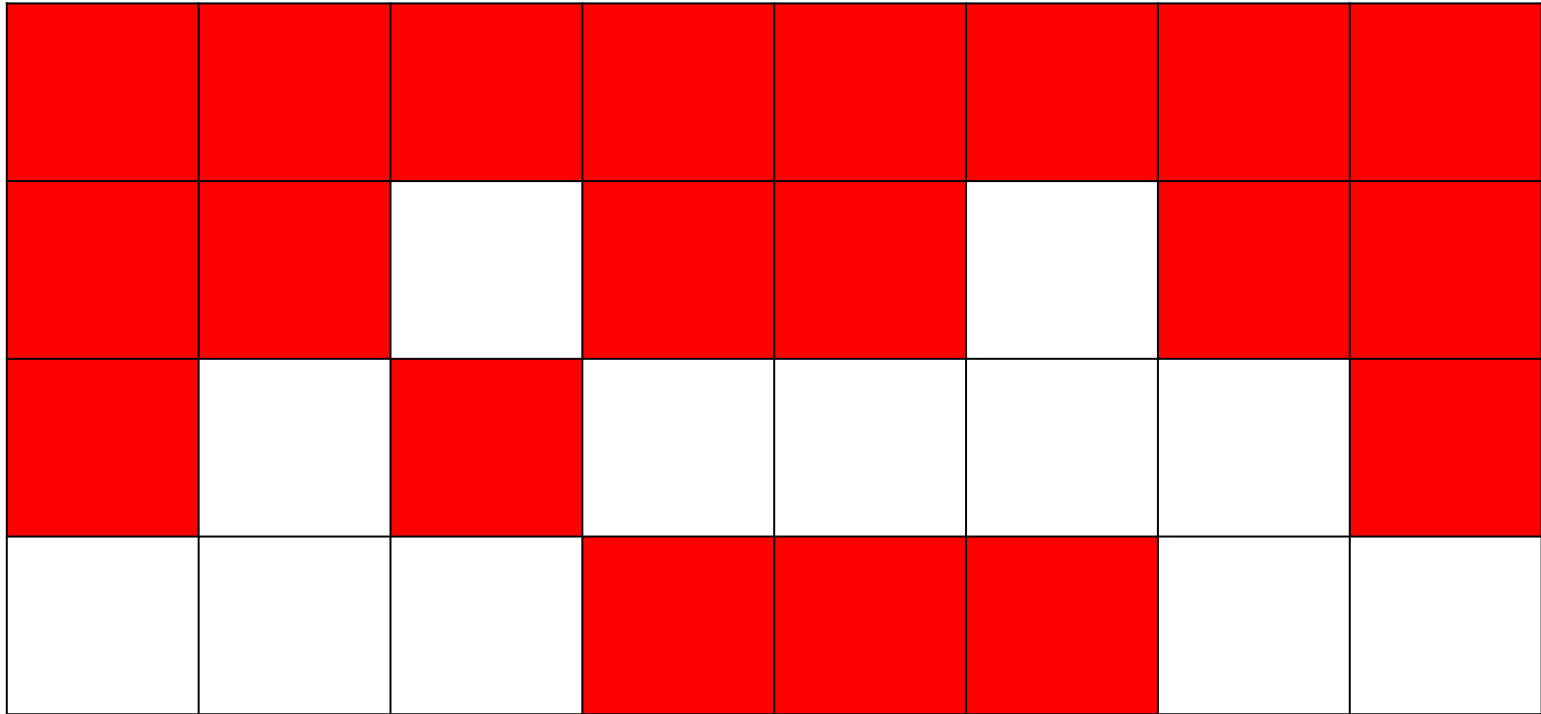
# Number of Classes



Play around with the number of classes & boundaries  
Or try a continuous choropleth map



# Aggregation



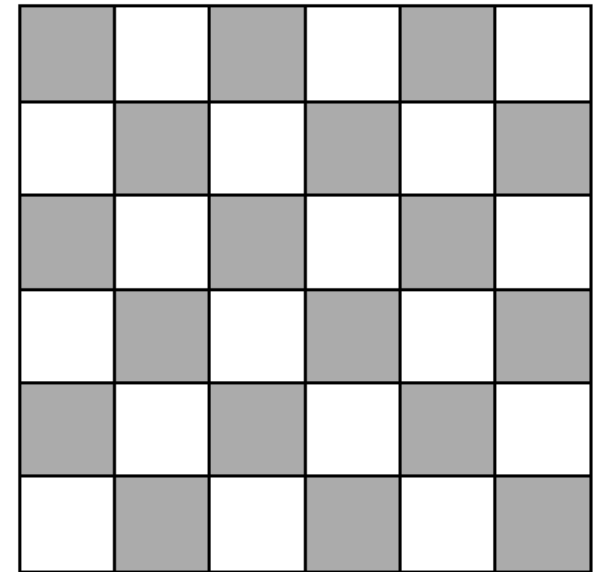
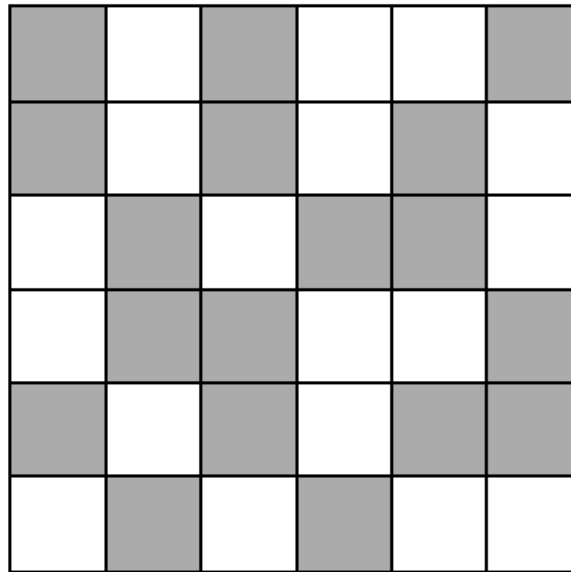
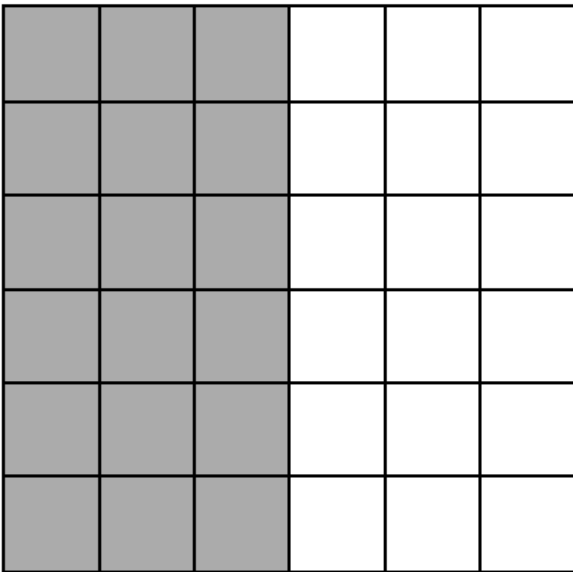
Aggregation can easily mislead

# Aggregation





# Spatial Autocorrelation



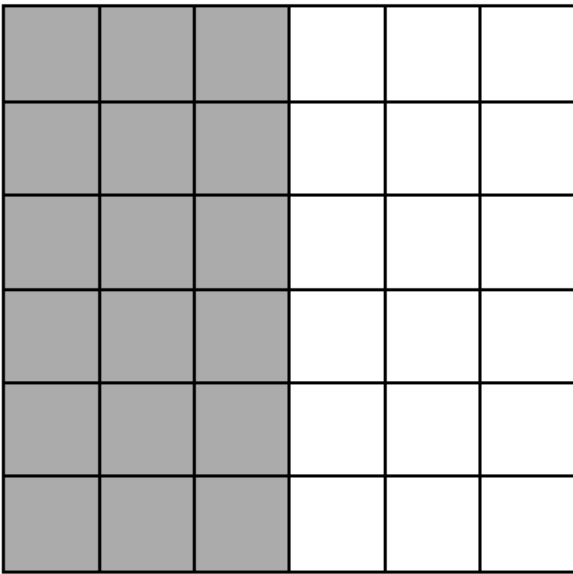
# Spatial Autocorrelation

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

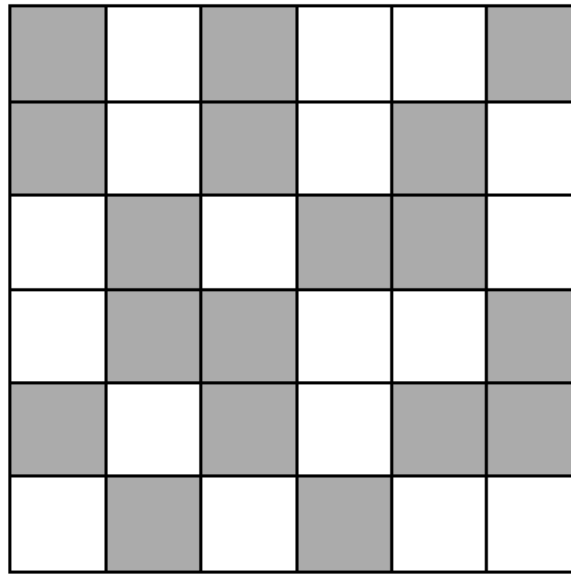
One way of measuring this is due to Moran  
(the Moran coefficient I)

- $w_{ij} = 1$  if contiguous, 0 o/w

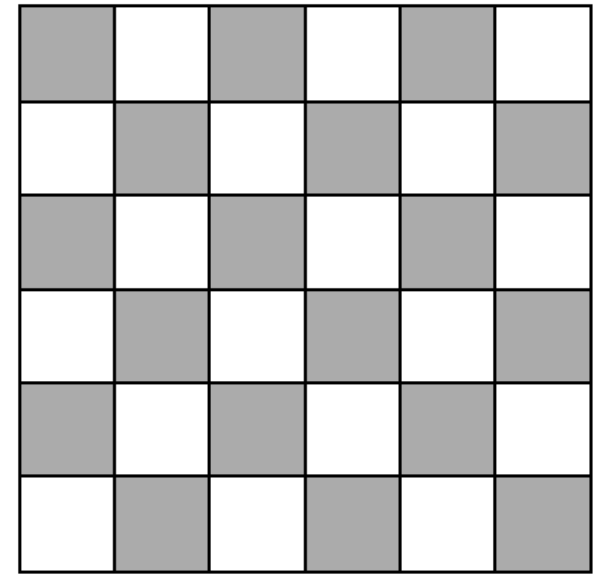
# Spatial Autocorrelation



Moran  $I \approx 1$



Moran  $I \approx 0$



Moran  $I \approx -1$

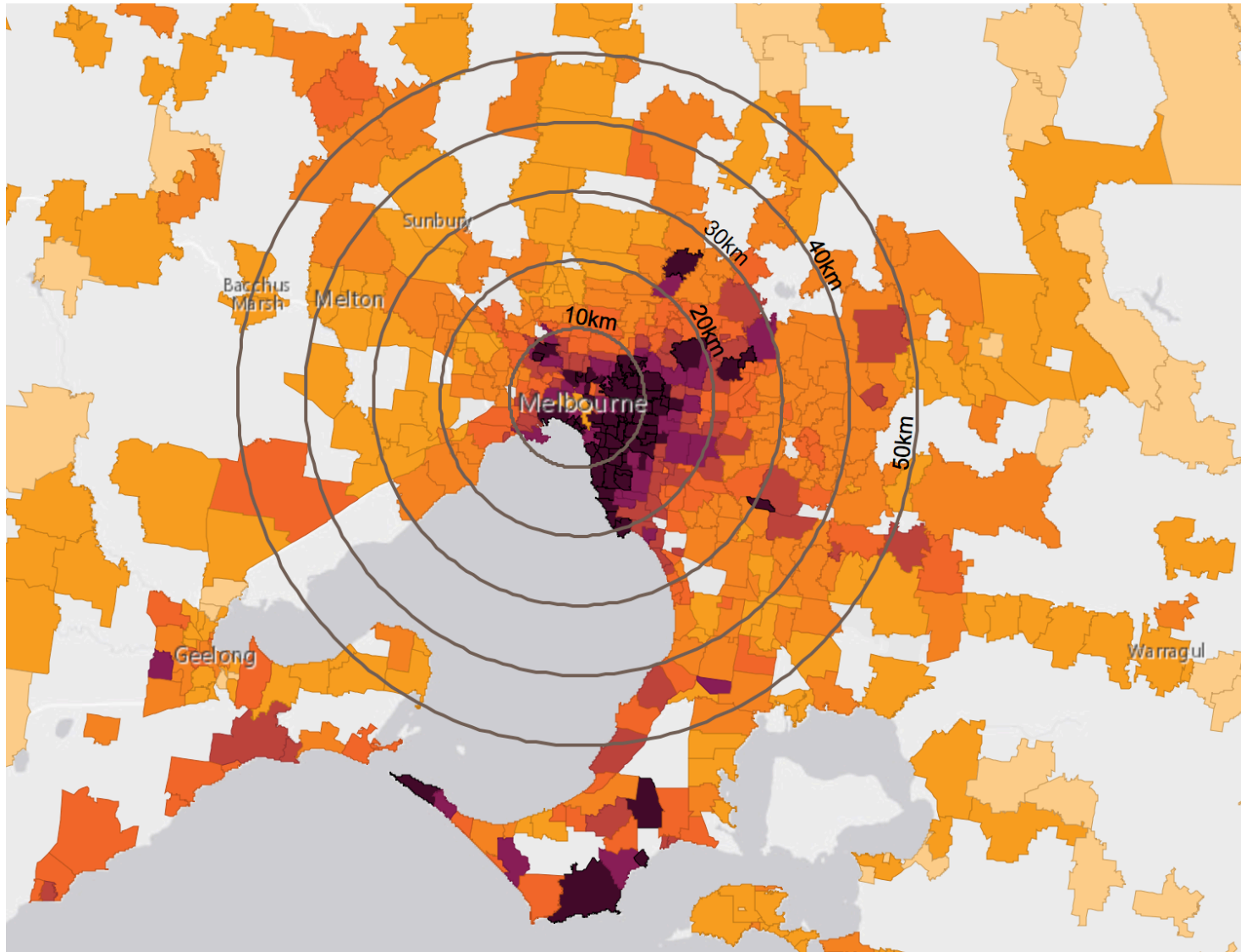
# Using Tabular Analysis

Spatial data can also be analysed using standard tabular visualisations and analyses

- Geographic distance can be used in clustering algorithms
- Contiguous areas can be grouped into regions which are treated as categories
- Latitude, longitude or distance can be treated as a quantitative variable



# Falling House Prices



## Melbourne

Houses	Units
In September 2013, 42.0% of suburbs in Melbourne had a median house value of less than \$500,000 compared to 3.8% by the end of September 2018.	The share of Melbourne suburbs with a median unit value of more than \$500,000 has increased to 67.9% in September 2018 from 23.0% in September 2013.
Current t	Current t
2013	2013

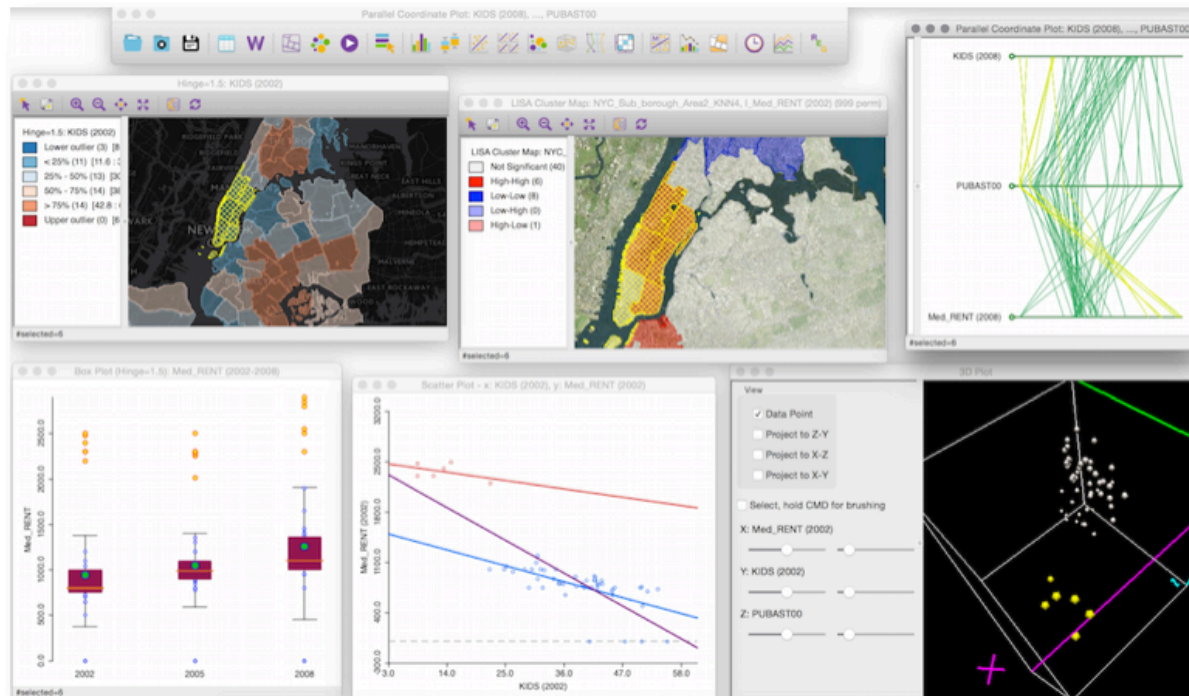
# Group Activity: Spatial Data Vis

How would you analyze the trends in house and unit prices in Melbourne? Questions are

- What are the main trends for different kinds of property?
- How does the location of the suburb affect the trend?
- How do features such as proximity to the sea, CBD, railway station affect the trend?

What visualisations and analyses would you use to explore these questions? You should consider a mix of spatial and tabular visualisations and analyses.

# Multivariate Spatial Data



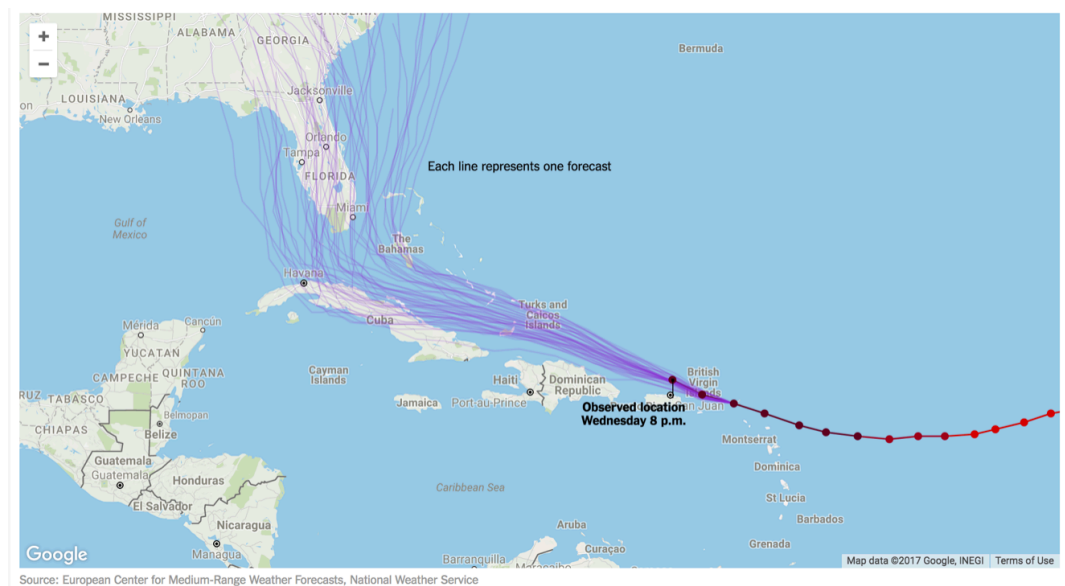
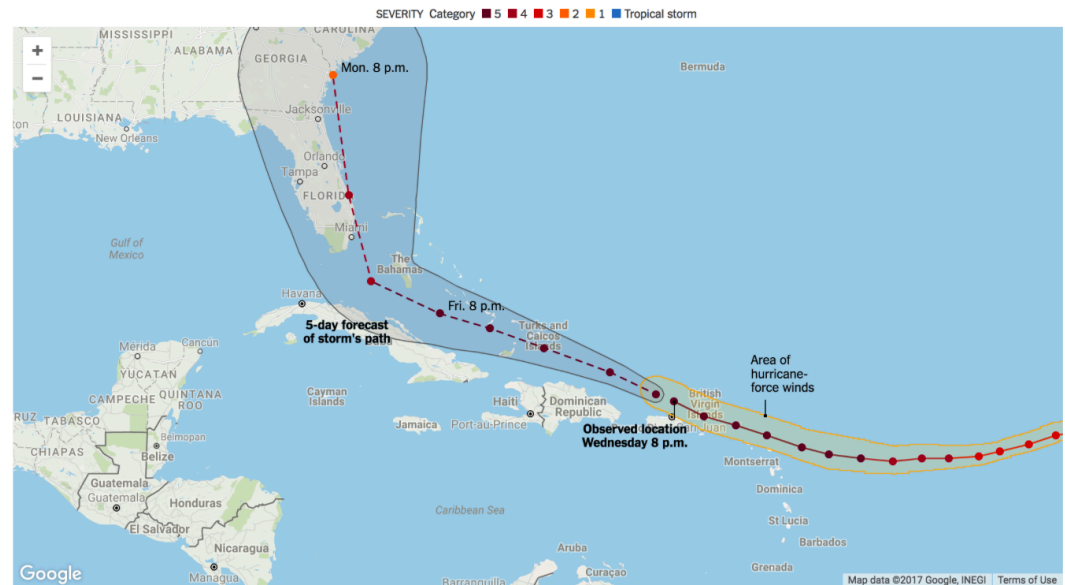
It is real challenge to show spatial data with multiple dimensions (x,y already taken)

- Small multiples
- Linked views

# Uncertainty

Showing uncertainty

- **Intrinsic:** combine uncertainty with attribute being visualised
- **Extrinsic:** separate depiction of uncertainty



# Announcements

## **Tutorial Activities**

MapBox; discuss project

## **2<sup>nd</sup> Online Quiz**

Opens next week; covers material from **Weeks 4 and 5**

## **Programming Exercise 2: R**

Due end of this week (5%)

## **Data exploration project**

Continue exploring your data

Less than 4 weeks until it is due (End Mid-Semester Break)