



FIT5147 Data Exploration & Visualisation

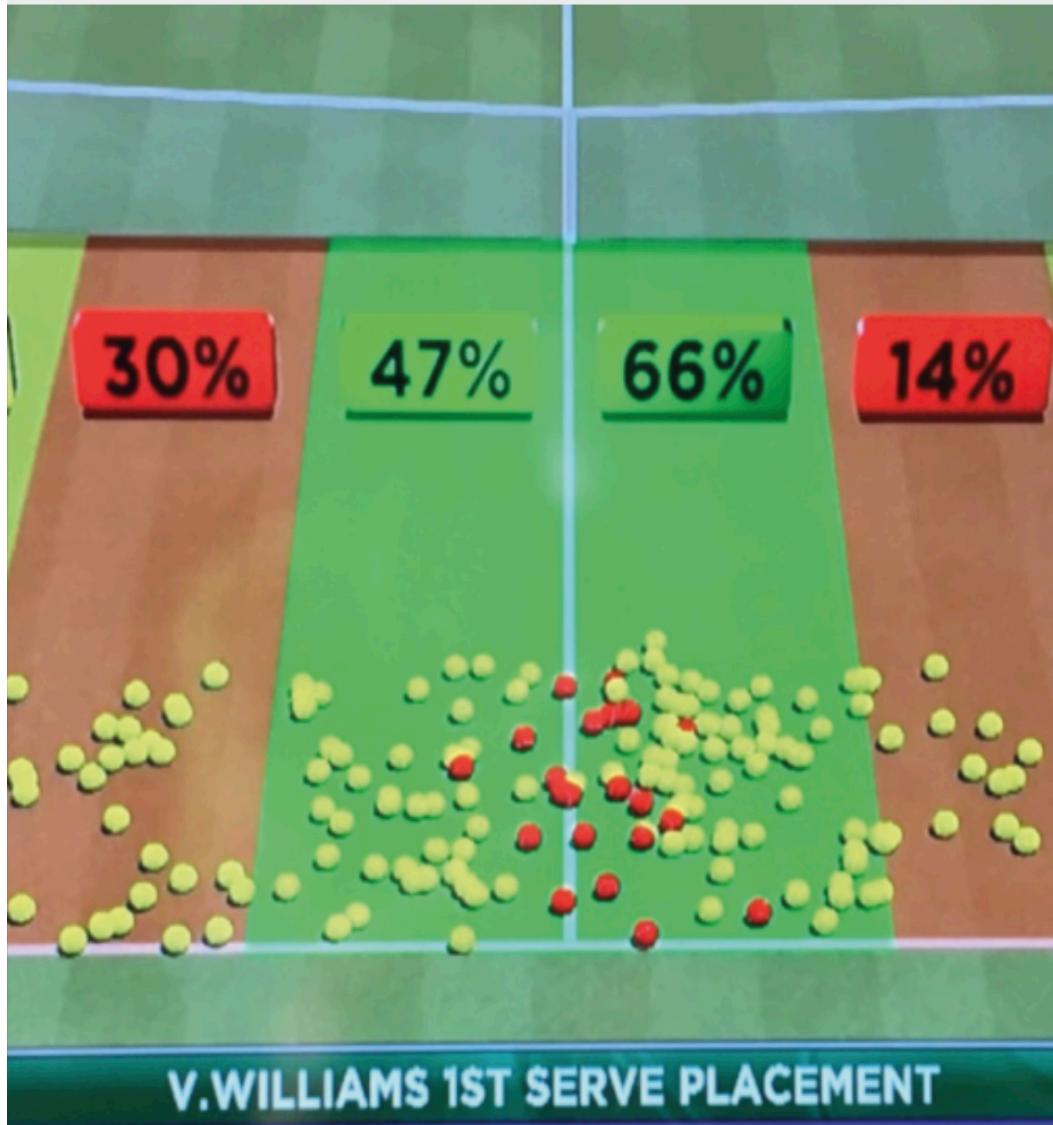
Kim Marriott

Hall of Fame/Shame Tue 2pm



Liang Liang (<https://projects.christianlaesser.com/travel-visa-inequality/>)

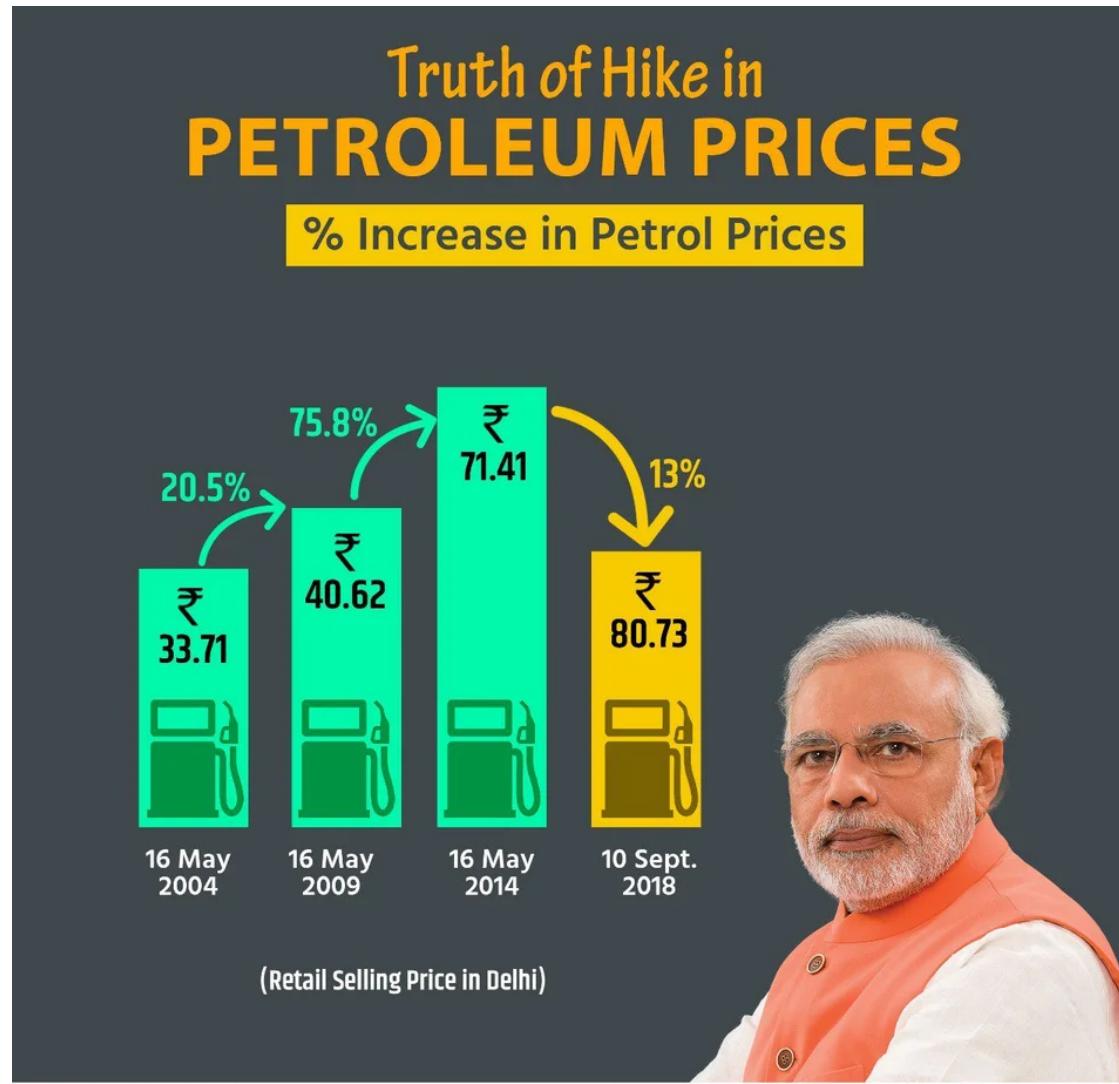
Hall of Fame/Shame Tue 2pm



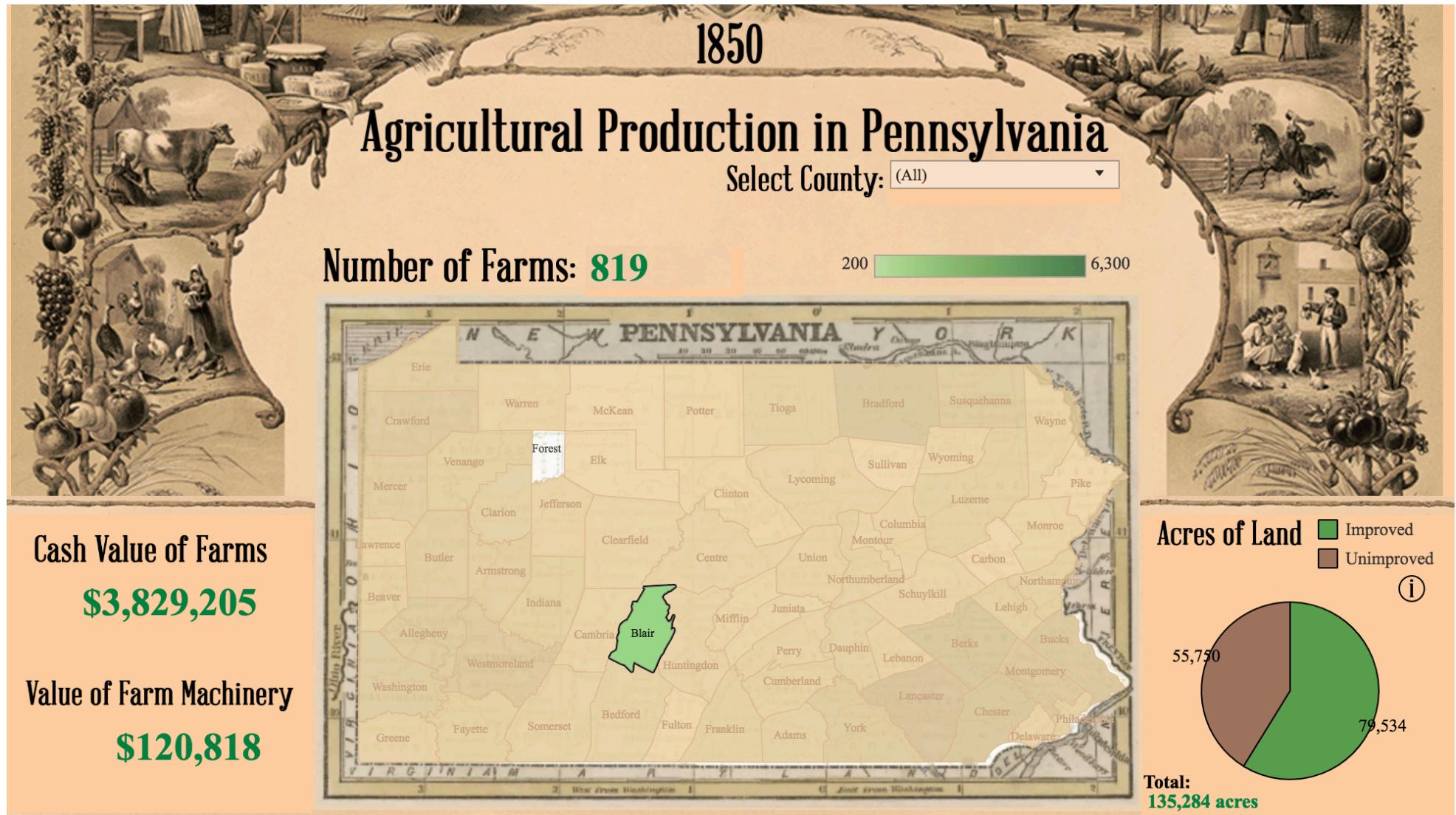
HsinHui Lin

(<http://www.visualisingdata.com/2017/07/little-visualisation-design-part-41/>)

Hall of Fame/Shame Wed I I am



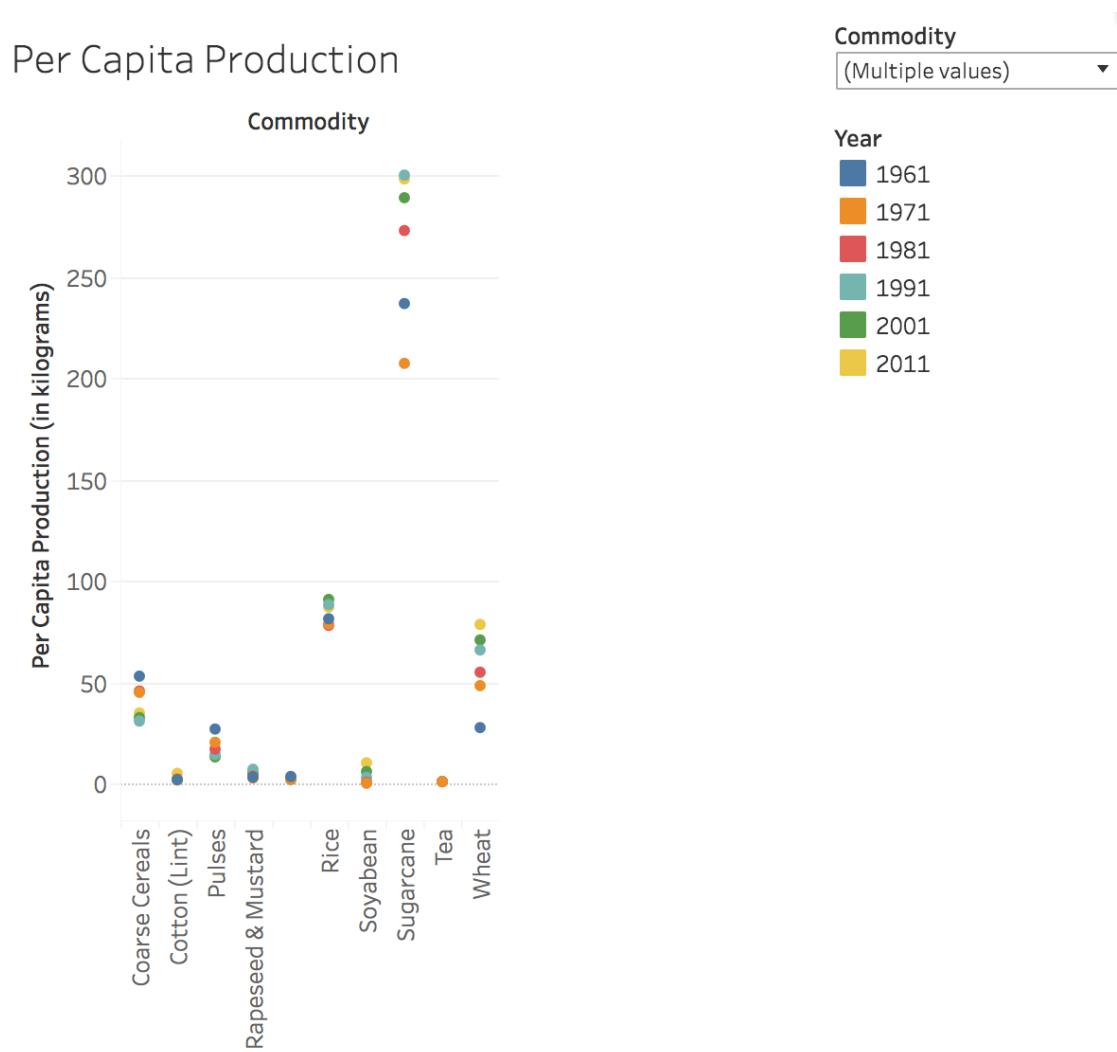
Hall of Fame/Shame Wed I I am



Cristiana Garcia Gewerc

(<https://public.tableau.com/profile/privatestorm#/vizhome/shared/94SCG5MCQ>)

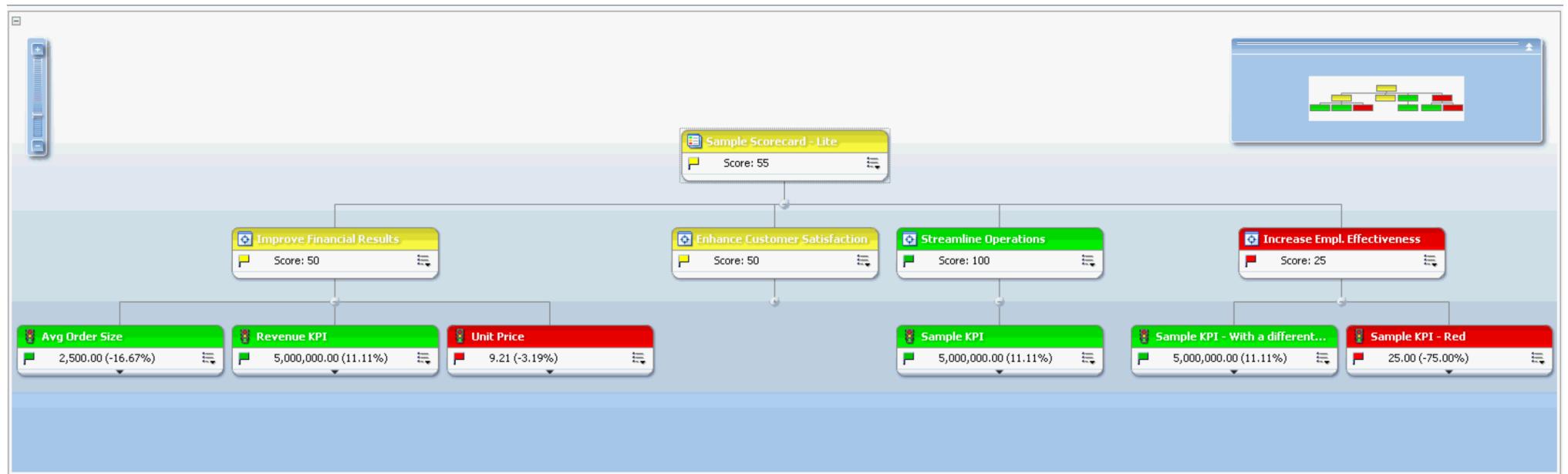
Hall of Fame/Shame Wed I Iam



Cristiana Garcia Gewerc

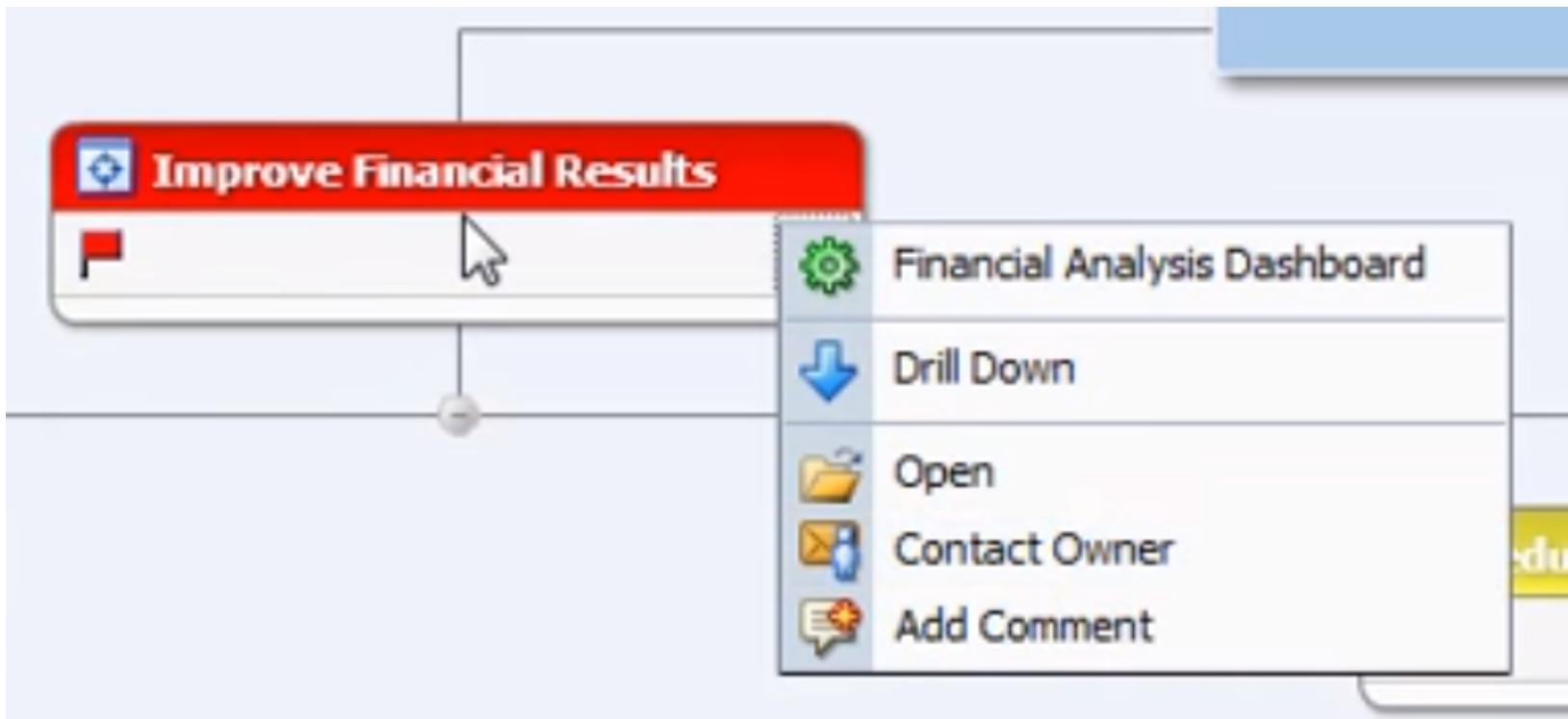
(https://public.tableau.com/shared/MY7MBXTX7?:display_count=yes)

Hall of Fame/Shame Wed 4pm



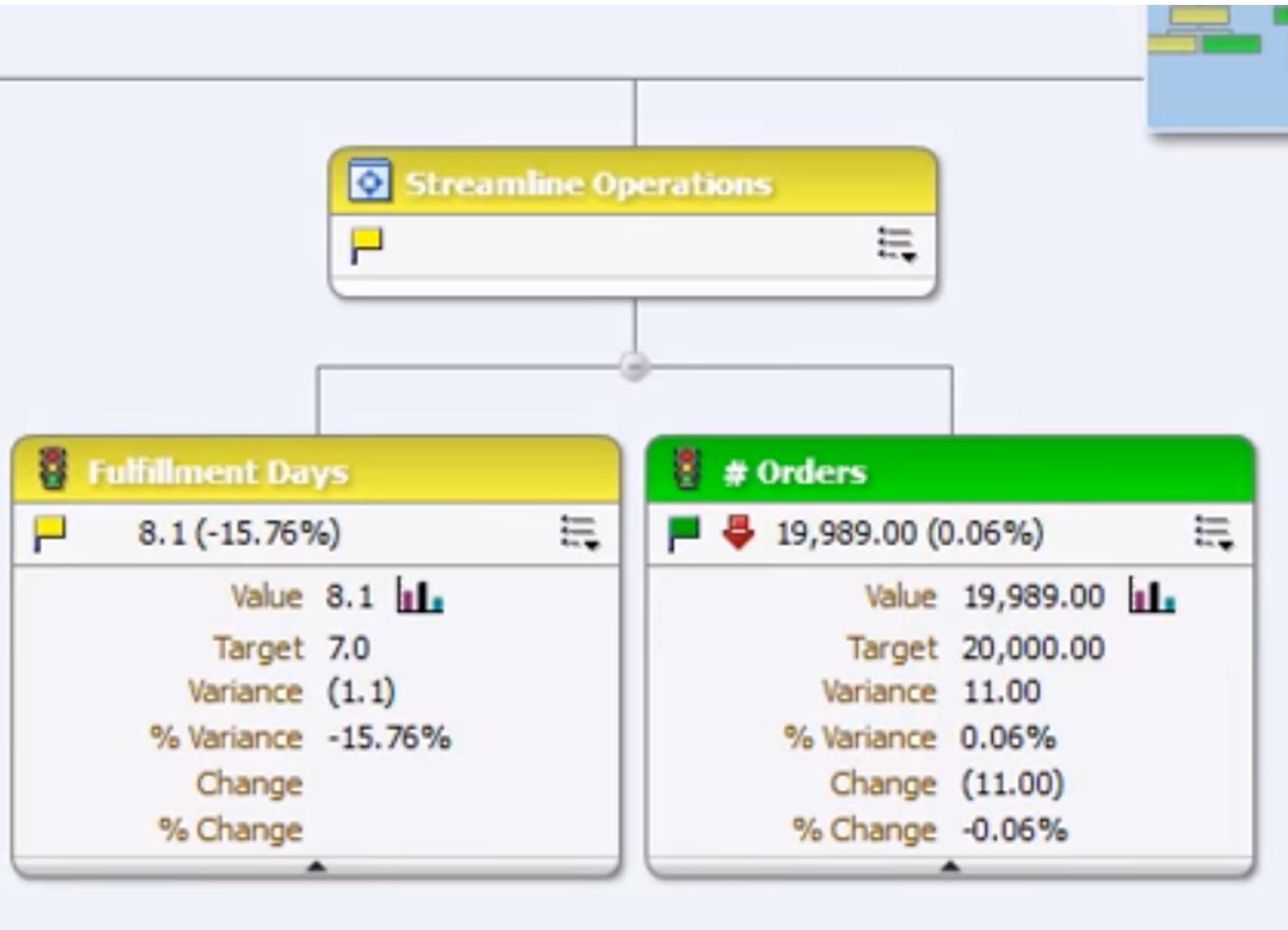
Tooba Jalalidil (Oracle BI)

Hall of Fame/Shame Wed 4pm



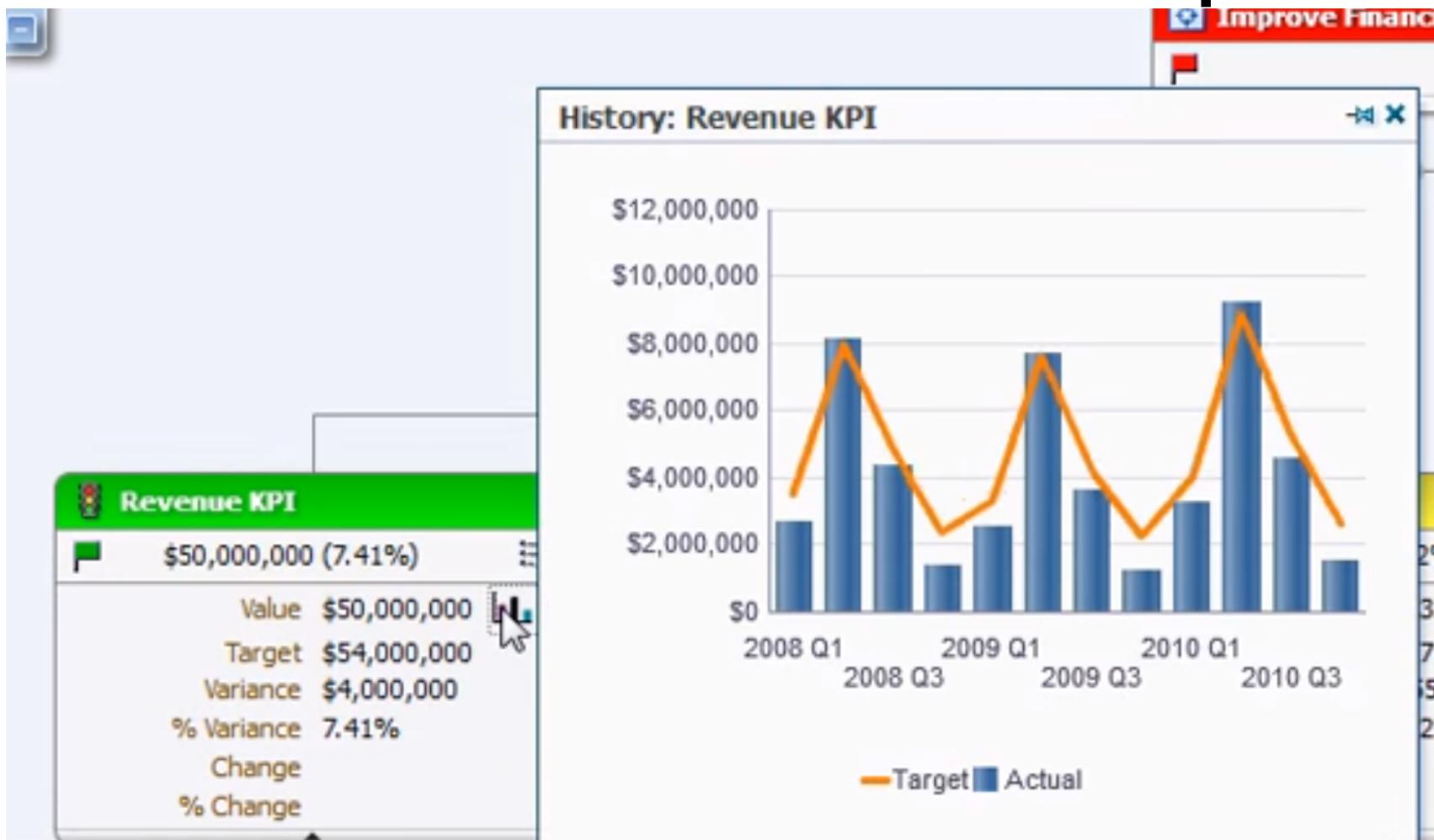
Tooba Jalalidil (Oracle BI)

Hall of Fame/Shame Wed 4pm



Tooba Jalalidil (Oracle BI)

Hall of Fame/Shame Wed 4pm



Tooba Jalalidil (Oracle BI)

Syllabus

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

Main Kinds of Data Sets

Tabular data: Data organised in tables, a row for each data item and a column for each of its attributes.

Spatial data: Data which is naturally organised and understood in terms of its spatial location or extent.

Network data: Nodes in the network are data items and links between the nodes are relations between. For instance a social network.

Textual data: this kind of data set consists of sequences of words and punctuation.

Visualising Tabular Data



Most statistical graphics
are for tabular data

Group Activity

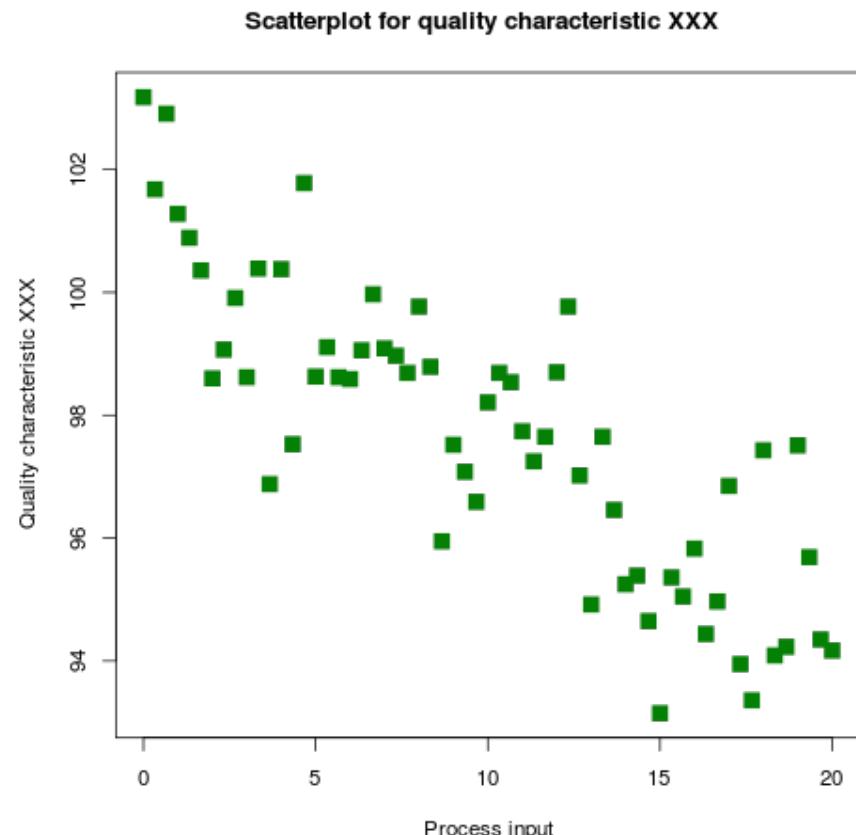
Working in groups of 5-6 people or by yourself if live streaming

- Name and draw at least 4 statistical graphics used to show the relationship between two variables? Which variable is usually dependent, which independent?
- Name and draw at least 3 statistical graphics used to show the distribution of a single variable?

Common Tabular Visualisations

Relationship between
two variables

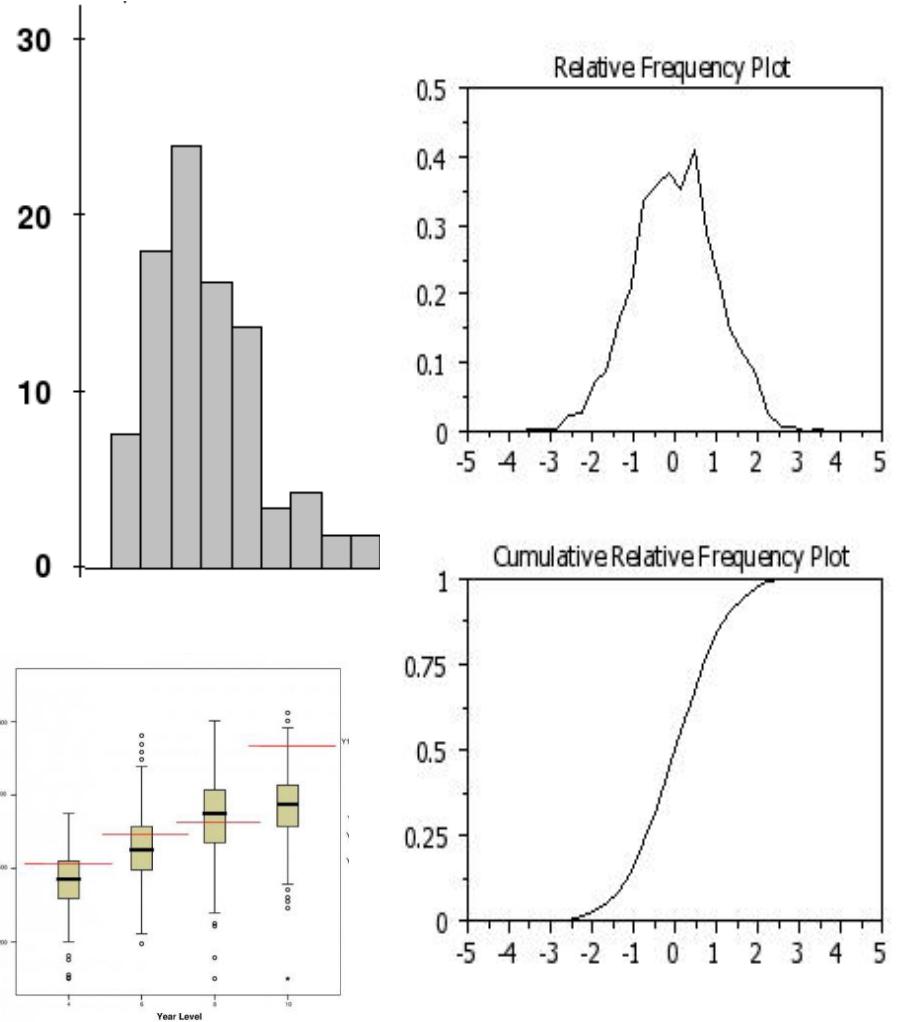
- Bar chart
- Line graph
- Pie chart
- Dot chart
- Scatter plot



Common Tabular Visualisations

Distribution of single variable

- Frequency or density plot
- Histogram
- Cumulative frequency diagram
- Box-and-whisker plot



Less Common Tabular Visualisations

Divide into groups of 5-6.

Each group should grab a piece of paper and pen.

Then create for the visualisation associated with the number on the piece of paper the

- Name(s) for that visualisation
- Example drawings of standard form and variants
- Purpose and what it is good for
- Closely related visualisations

If you are watching on live-streaming please choose one that you haven't heard of and do the same thing by yourself.

Statistical Graphics

- | | | |
|-----------------------------------|---------------------------------|----------------------------|
| I. Paired bar chart | 10. Gantt chart | 20. Violin plot |
| 2. QQ Plot | 11. Mosaic plot | 21. Hexbin plot |
| 3. Parallel coordinates | 12. SPLOM (Scatter plot matrix) | 22. Polar area diagram |
| 4. Bullet graph | 13. Area chart | 23. Bubble chart |
| 5. Heat map (table) | 14. Stream graph | 24. Horizon chart |
| 6. Stem-and-leaf plot | 15. Wind rose | 25. Connected scatter plot |
| 7. Spider chart (aka radar chart) | 16. Tree map | |
| 8. Stacked bar chart | 17. Doughnut chart | |
| 9. Compound bar chart | 18. Chernoff faces | |
| | 19. Small multiples | |

Multivariate Visualisation

Most real-world data is ***multi-dimensional***

One of the big challenges in data visualisation is dealing with multi-dimensional data (escaping flatland – Tufte)

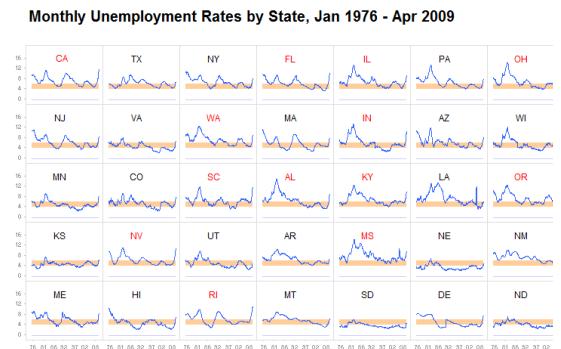
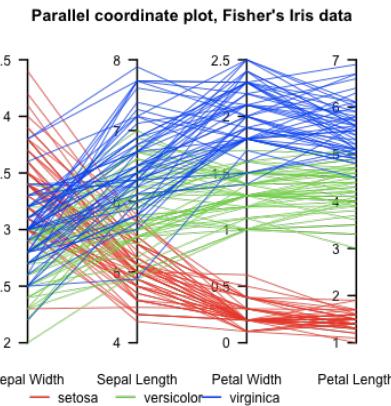
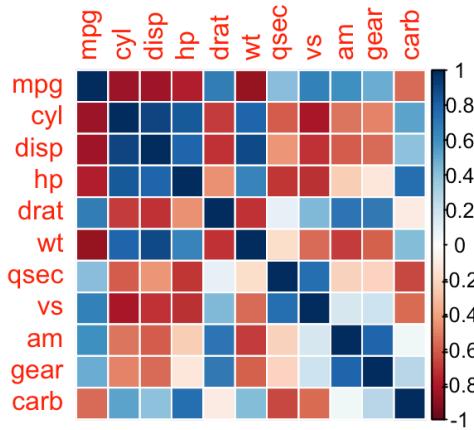
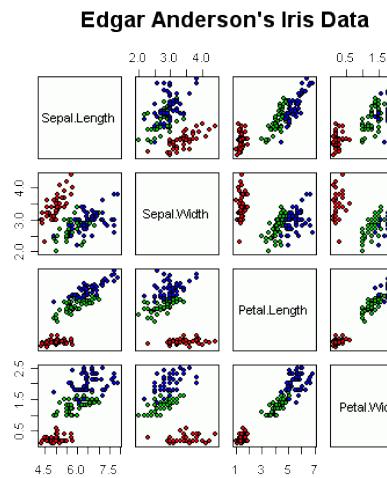
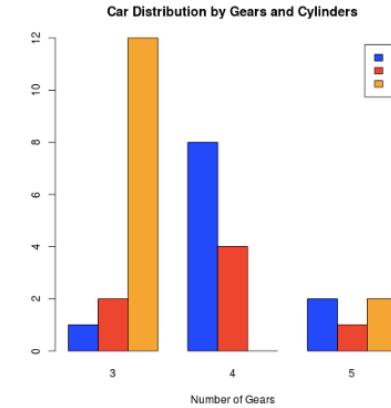
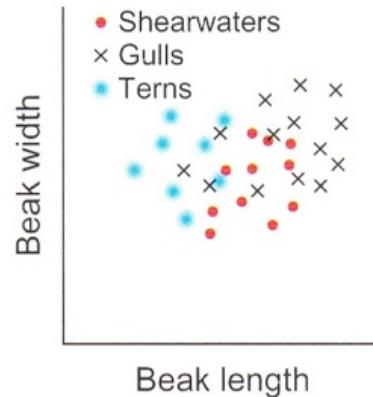
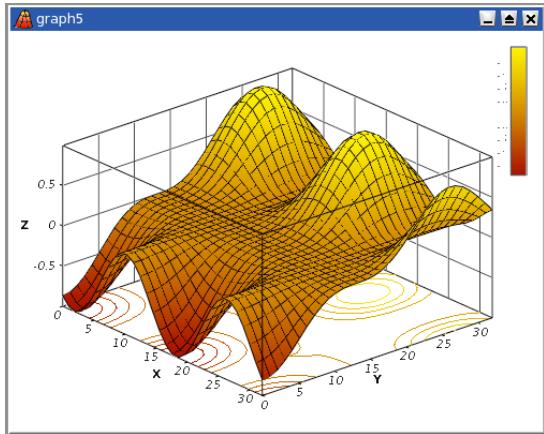
Called **multivariate** visualisation

What are some useful multivariate visualisations for multi-dimensional data?

Statistical Graphics

- | | | |
|-----------------------------------|---------------------------------|----------------------------|
| I. Paired bar chart | 10. Gantt chart | 20. Violin plot |
| 2. QQ Plot | 11. Mosaic plot | 21. Hexbin plot |
| 3. Parallel coordinates | 12. SPLOM (Scatter plot matrix) | 22. Polar area diagram |
| 4. Bullet graph | 13. Area chart | 23. Bubble chart |
| 5. Heat map (table) | 14. Stream graph | 24. Horizon chart |
| 6. Stem-and-leaf plot | 15. Wind rose | 25. Connected scatter plot |
| 7. Spider chart (aka radar chart) | 16. Tree map | |
| 8. Stacked bar chart | 17. Doughnut chart | |
| 9. Compound bar chart | 18. Chernoff faces | |
| | 19. Small multiples | |

Multivariate Visualisation



Source: Bureau of Labor Statistics

Notes: The orange band denotes a "normal" unemployment rate (4%-6%). State code in red: unemployment rate in April 2009 is higher than the US average

Visual Language

Visualisations work by mapping data items and attributes to spatial and visual attributes of geometric shapes

Barchart

- Each item maps to an axis-aligned rectangle
- Key (independent) attribute maps to x-position of rectangle.
- Value (dependent) attribute maps to height of rectangle

Visual Language

These mappings are called **visualisation idioms**

Gives a finite set of rules for visualisation

Can be creative in how you use them but constrained in what you can do

A kind of **visual language**

Jacques Bertin: *Semiologie Graphique* (Semiology of Graphics) 1967

Graphics in R

Lots of graphics packages in R.

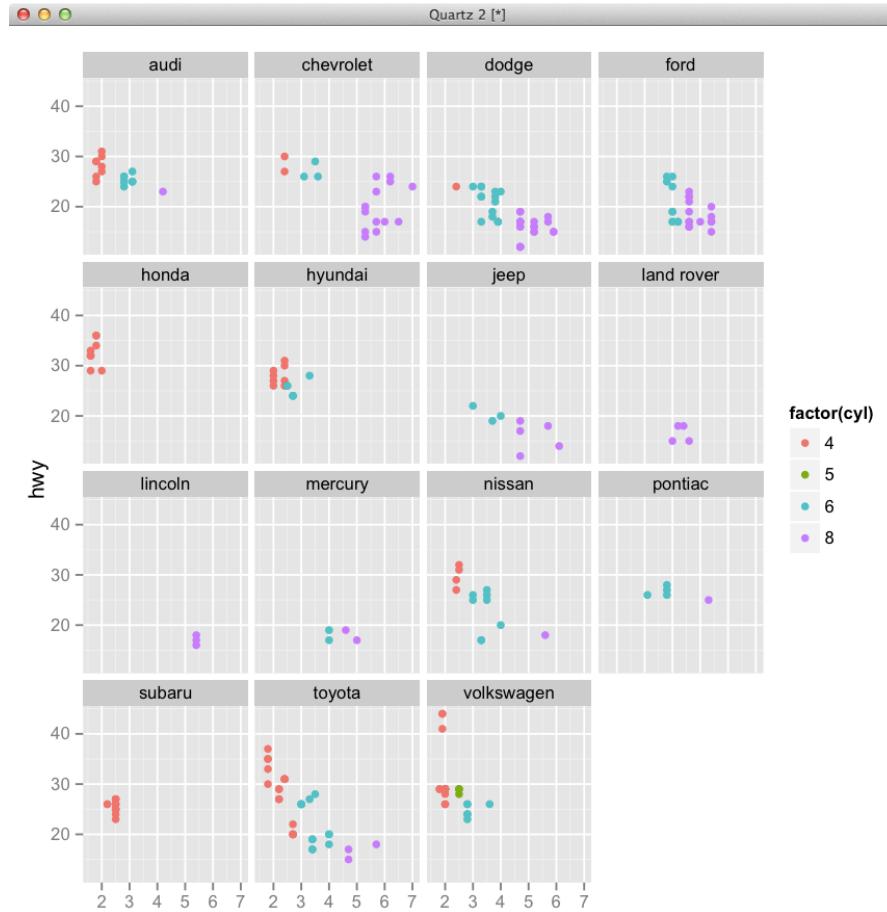
ggplot2 (by Hadley Wickham) is one of the most widely used

It is based on a formal grammar of statistical graphics

Very, very expressive

Activity in this week's tute

ggplot2



ggplot2 is based on a layered model of a statistical graphics

- A graphic is composed of *facets*
- A facet is composed of *layers* and a *scale*

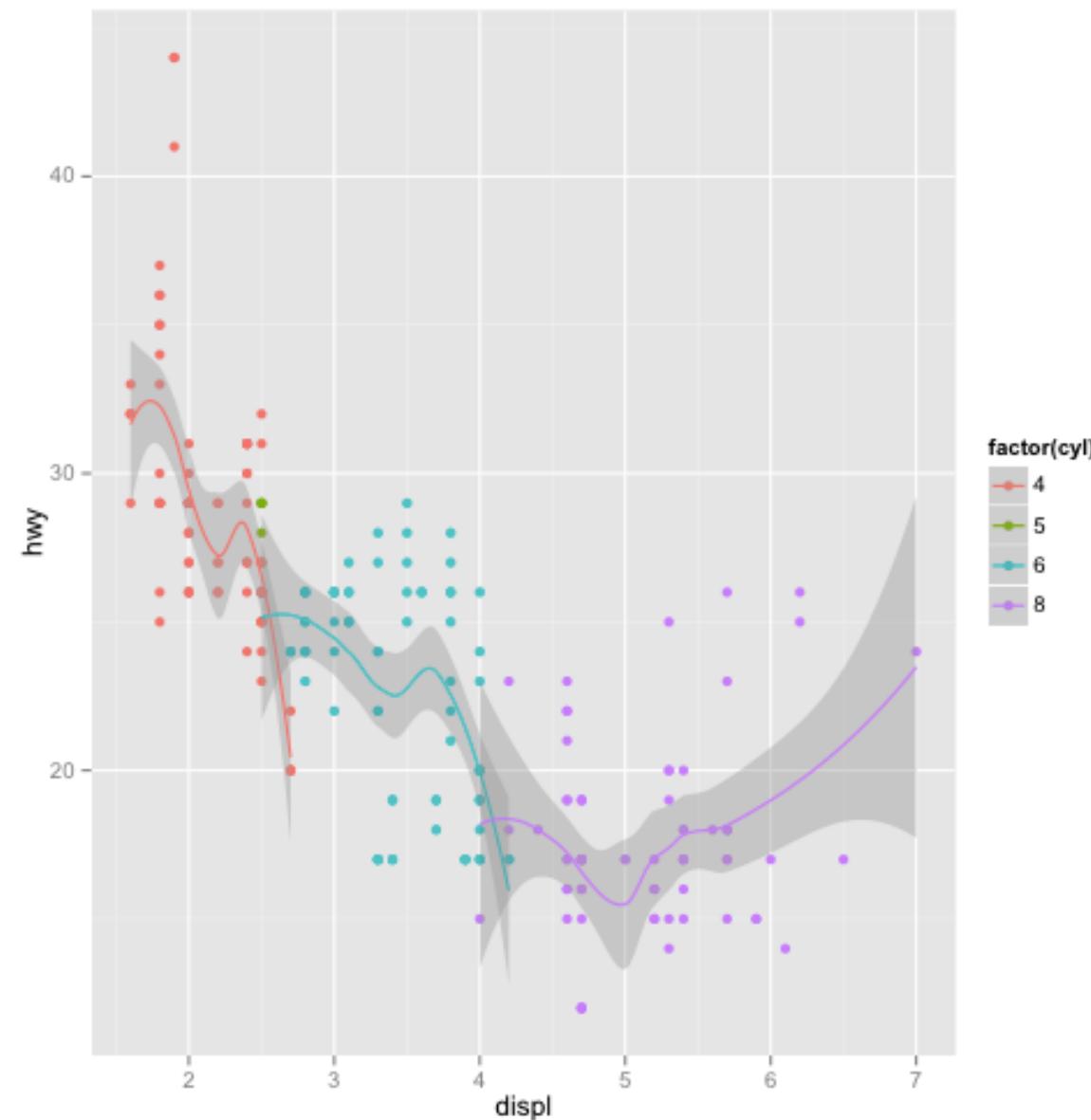
Layers

A *layer* is composed of

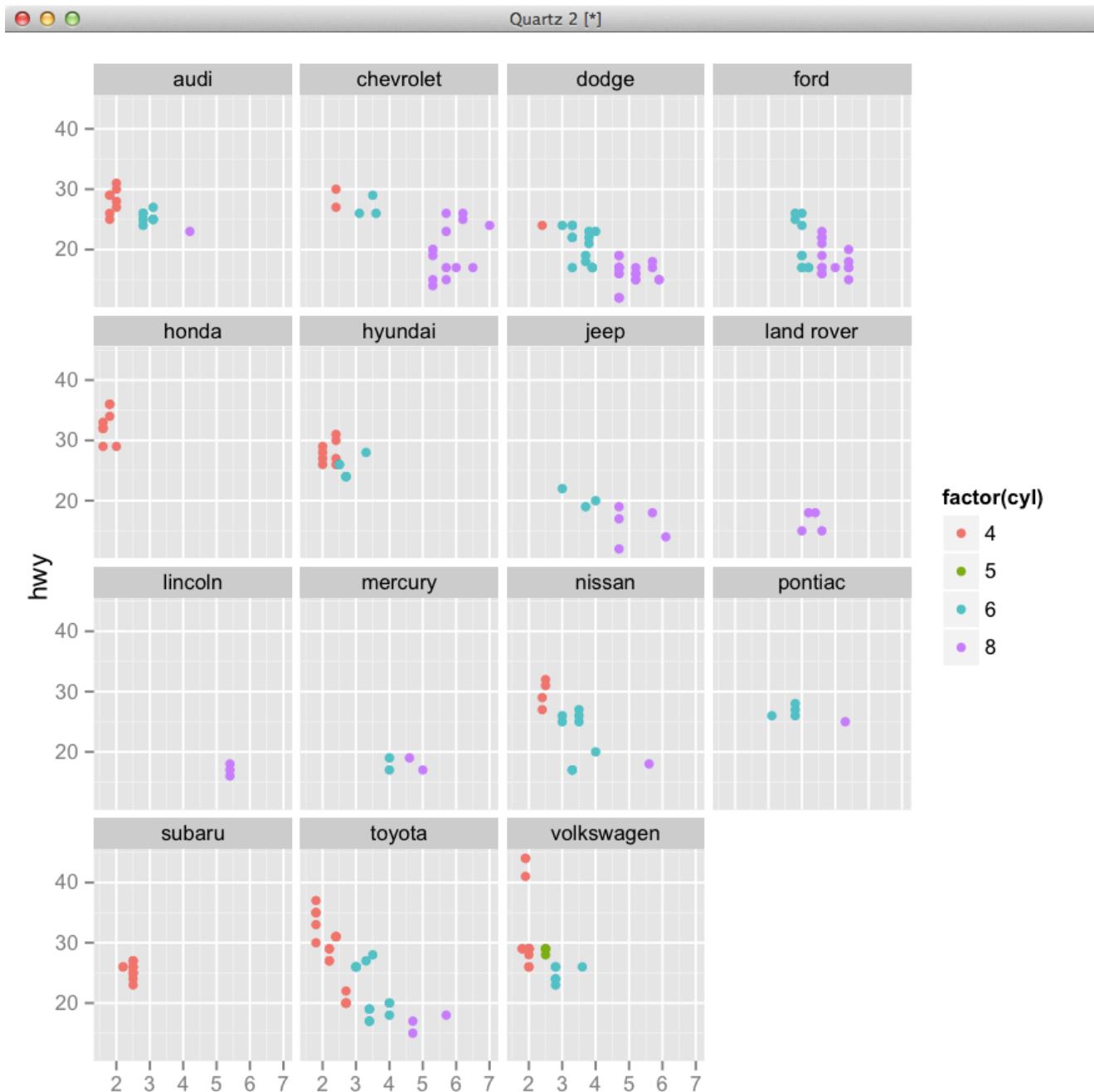
- *Data*: which must be a data frame
- *Aesthetics*: map data to a visual representation
- *Geoms*: gives the geometric representations
- *Stat*: Statistical transformation to apply to data
- *Position adjustment*: how to adjust position to avoid over plotting

○ ○ ○

Quartz 2 [*]



```
qplot(displ, hwy, data=mpg, colour=factor(cyl), geom=c("point","smooth"))
```



```
qplot(displ, hwy, data=mpg, colour=factor(cyl)) +facet_wrap(~manufacturer)
```

Assessment

Programming Exercise I: Tableau Public Due end of the week (5%)

Data Exploration Project Proposal Email to your tutor THIS WEEK for approval.

PLEASE DON'T START UNTIL YOU GET EMAIL APPROVAL