



EMPIRICAL POLITICAL RESEARCH

Presentation of Data/ Tutorial 6 by Aleksandra Butneva

WHY DO WE NEED GRAPHS?

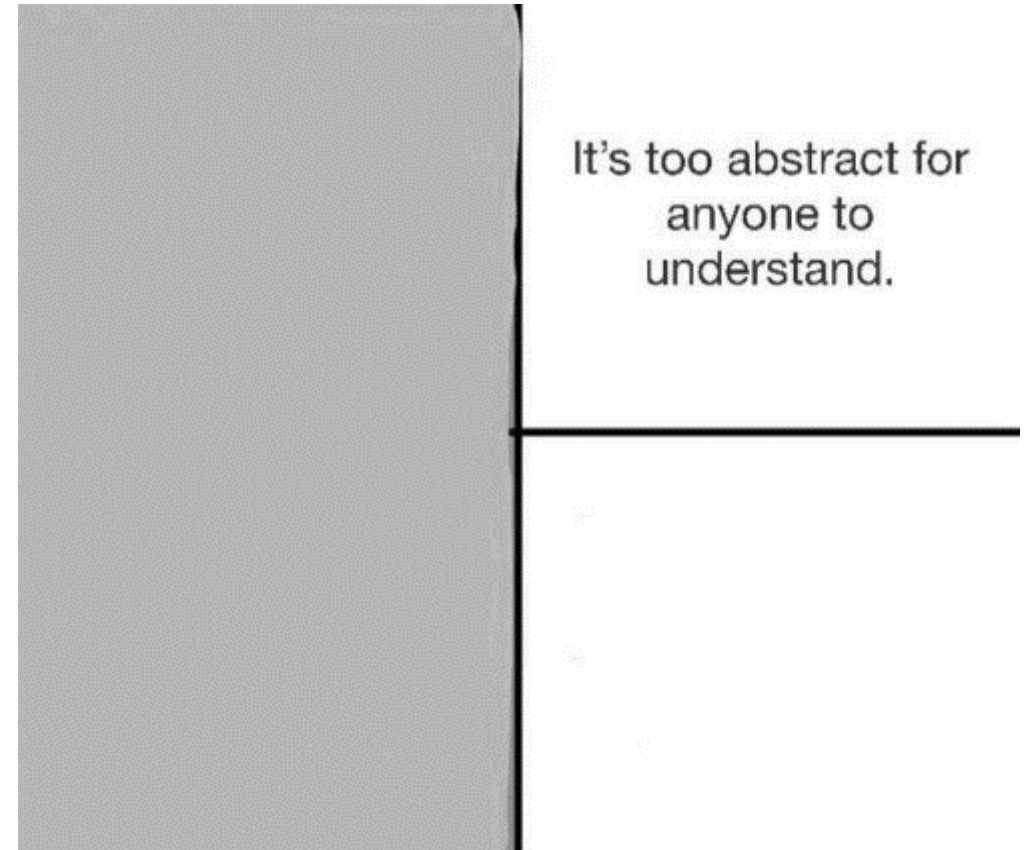
Main purpose: communicate your findings
(=scientific knowledge gained throughout your research)

Simplify - our theories and the complexity behind the data

Compare - can make comparisons across variables, space or time

Identify trends - we can tell a story if we use the right visualisation tool

- however, graphs could be misused (which happens quite often)



Can they understand?

HOW DO YOU GET THERE? EXPERT OPINIONS ABOUT GOOD GRAPHS GATHERED...

Draw only necessary lines and shapes

Avoid redundancy: do not repeat what has already been said

Work in black-and-white (you have to pay for colours in publications)

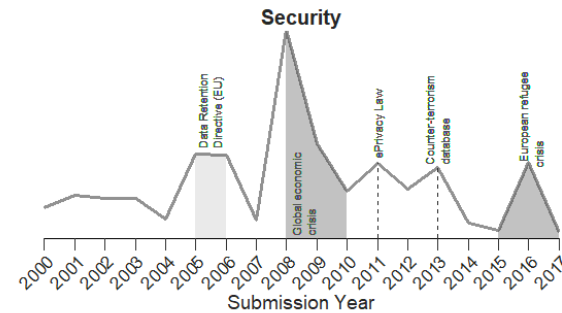
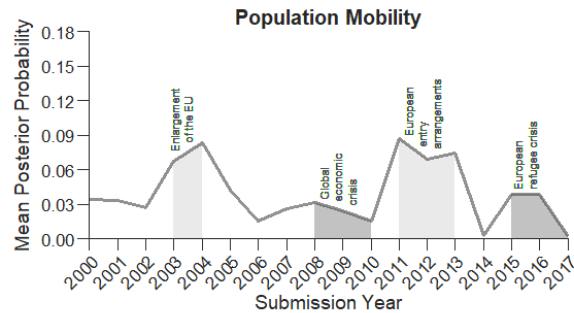
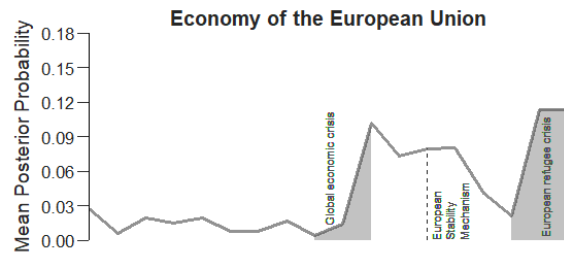
No chart-junk; 3D-effects are dangerous

Be intuitive and consise, get your message across



HOW ARE GRAPHS BORN?

What you see:



What I see:

```
342 # Plot mean theta for Security
343 p4<-ggplot(means, aes(x=as.numeric(Year), y=`Security`)) +
344
345   geom_line(color = "grey15", size=1.1, alpha=0.5) +
346   geom_area(data = means %>% filter(Year > 2007 & Year < 2011), alpha = 0.3) +
347   geom_area(data = means %>% filter(Year > 2014), alpha = 0.3) +
348   geom_area(data = means %>% filter(Year > 2004 & Year < 2007), alpha = 0.1) +
349   theme_tufte(base_family = "Arial") +
350   ylab("") +
351   xlab("Submission Year")+
352   labs(title = "Security") +
353   scale_x_continuous(breaks = seq(2000, 2017, by = 1)) +
354   scale_y_continuous(breaks = seq(0, 0.18, by = .03)) +
355   coord_cartesian(xlim = c(2000, 2017), expand = F) +
356   coord_cartesian(ylim = c(0, 0.18), expand = F) +
357   theme(axis.text.x = element_text(
358     size=13, angle=45, hjust = 1, color = "grey15"),
359     axis.title = element_text(size = 14, color = "grey15")) +
360   theme(axis.text.y = element_text(size=13, color = "grey15")) +
361   geom_text(aes(label = "Global economic"), y=0.003, x = 2008.3, angle = 90, hjust = 1,
362     family="Palatino Linotype", fontface="plain", color = "grey32")+
363   geom_text(aes(label = "crisis"), y=0.003, x = 2008.8, angle = 90, hjust = 0,
364     family="Palatino Linotype", fontface="plain", color = "grey32")+
365
366   geom_text(aes(label = "European refugee"), y=0.067, x = 2015.8, angle = 90,
367     family="Palatino Linotype", fontface="plain", color = "grey32")+
368   geom_text(aes(label = "crisis"), y=0.062, x = 2016.3, angle = 90, hjust = 0,
369     family="Palatino Linotype", fontface="plain", color = "grey32")+
370   geom_text(aes(label = "Counter-terrorism"), y=0.065, x = 2012.7, angle = 90,
371     family="Palatino Linotype", fontface="plain", color = "grey32")+
372   geom_text(aes(label = "database"), y=0.065, x = 2013.2, angle = 90, hjust = 0,
373     family="Palatino Linotype", fontface="plain", color = "grey32")+
374   geom_text(aes(label = "Data Retention "), y=0.079, x = 2005.2, angle = 90, hjust = 1,
375     family="Palatino Linotype", fontface="plain", color = "grey32")+
376   geom_text(aes(label = "ePrivacy Law"), y=0.07, x = 2011,
377     . angle = 90, hjust = 0, size = 3,
```

GUIDELINES FOR CREATING PUBLISHABLE GRAPHS (NOT FUNNY GRAPHS)

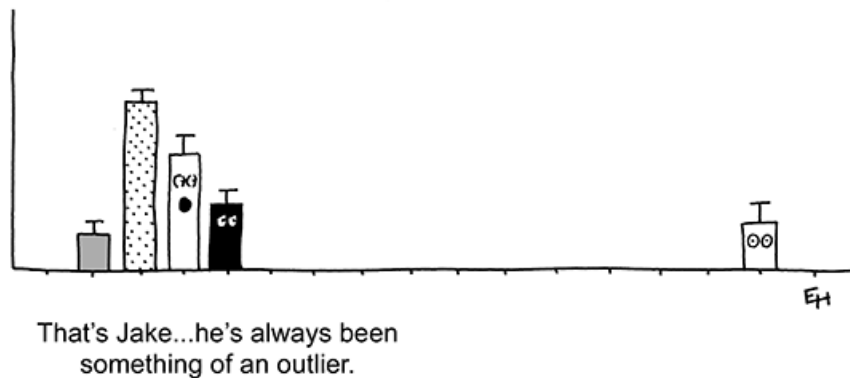
Data content: it does not make sense to use graphs to display very small amounts of data.

Relevance: graphs are only as good as the data they display

Complexity: avoid unnecessary complexity

Distortion: graph should not distort the values it portrays

Clear labelling: titles, axes, legends, data labels, etc.



GRAPH TYPES: PIE CHARTS

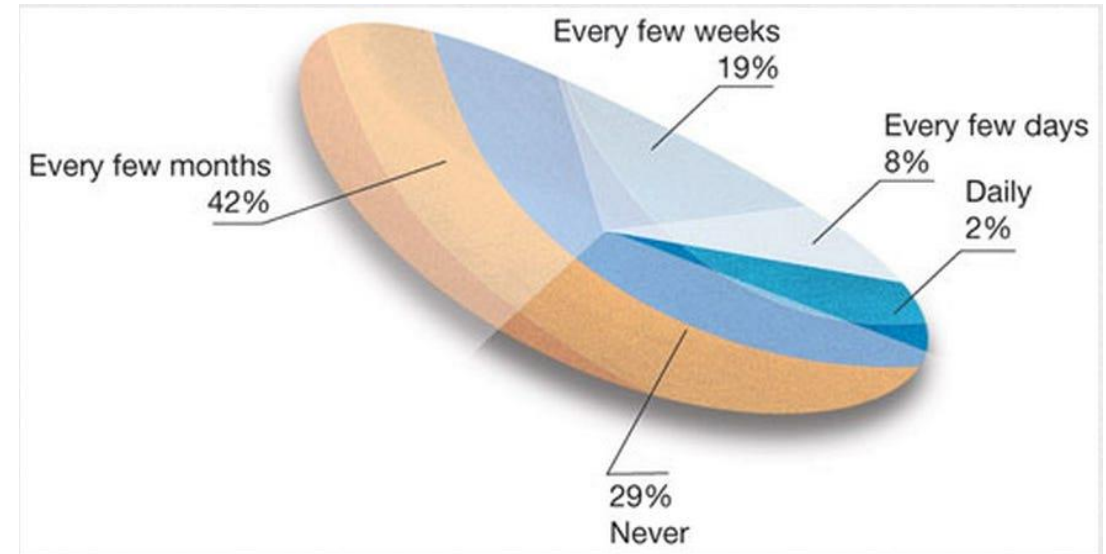
Presenting small amounts of data

Values need to sum to 100 %

You need to have variation between values

Rarely effective

What is wrong with this one?



Read more here:

<https://www.businessinsider.com/the-27-worst-charts-of-all-time-2013-6?r=DE&IR=T#but-hey-were-bipartisan-here-heres-a-chart-from-president-obamas-2013-state-of-the-union-address-that-isnt-a-venn-diagram-but-certainly-tries-to-look-like-one-it-also-compares-apples-and-oranges--china-and-the-us-have-wildly-different-energy-needs--while-not-exactly-drawing-the-contrast-or-comparison-that-the-president-actually-wanted-really-bad-chart-14>

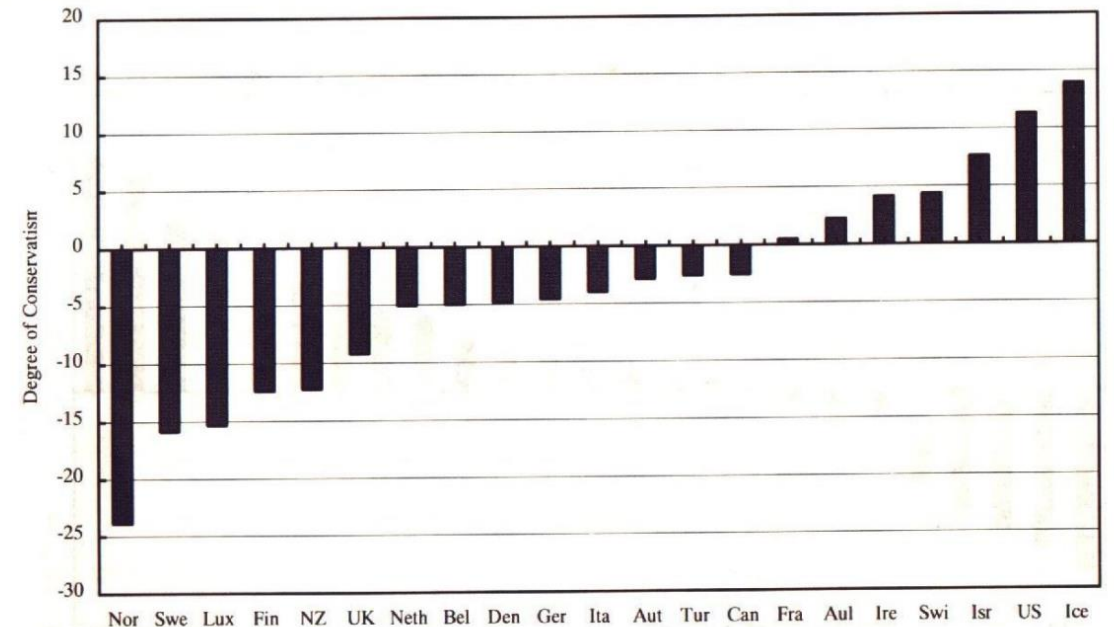
GRAPH TYPES: BAR CHARTS

When you need to visualize the frequency of your data

Used for nominal and ordinal variables

You can compare across quantities

What is wrong with this one?



Read more here:

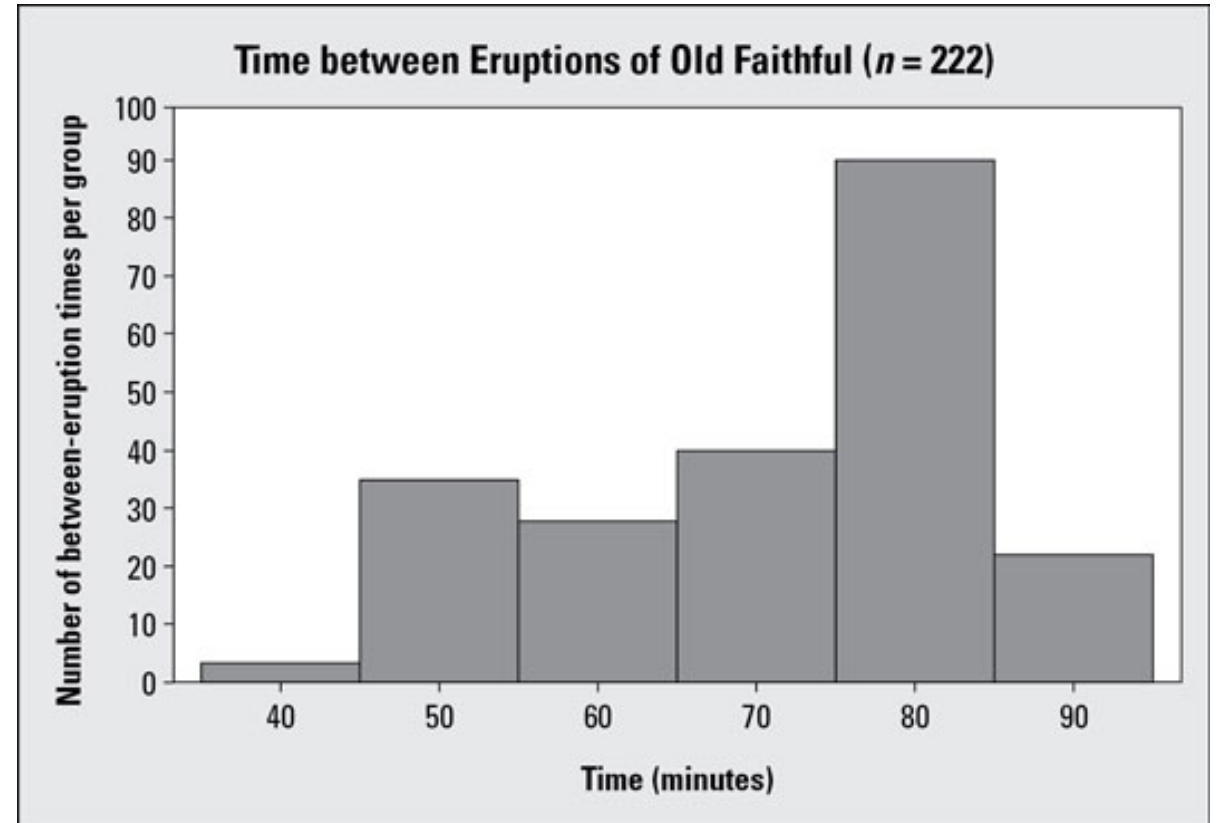
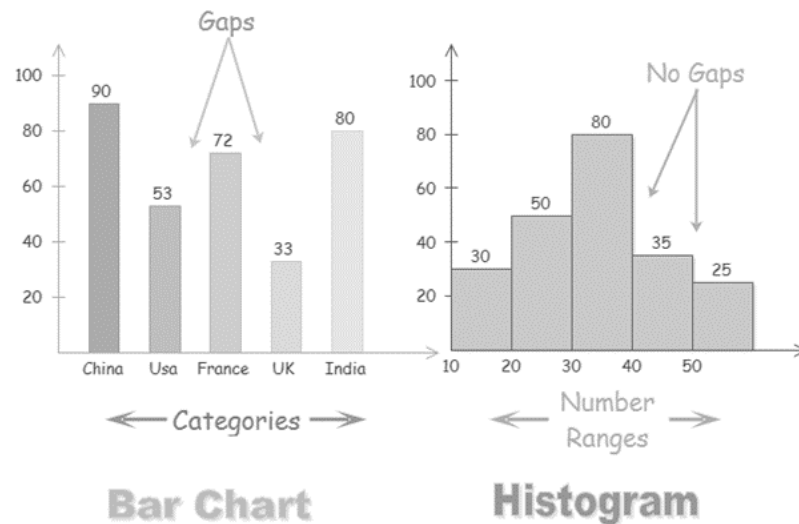
<https://www.businessinsider.com/the-27-worst-charts-of-all-time-2013-6?r=DE&IR=T#but-hey-were-bipartisan-here-heres-a-chart-from-president-obamas-2013-state-of-the-union-address-that-isnt-a-venn-diagram-but-certainly-tries-to-look-like-one-it-also-compares-apples-and-oranges--china-and-the-us-have-wildly-different-energy-needs--while-not-exactly-drawing-the-contrast-or-comparison-that-the-president-actually-wanted-really-bad-chart-14>

GRAPH TYPES: HISTOGRAMS

What is wrong with this one?

Show distribution of our variables

**Range needs to be divided in
equally spaced intervals, number
of cases are counted in bins**



Read more here:

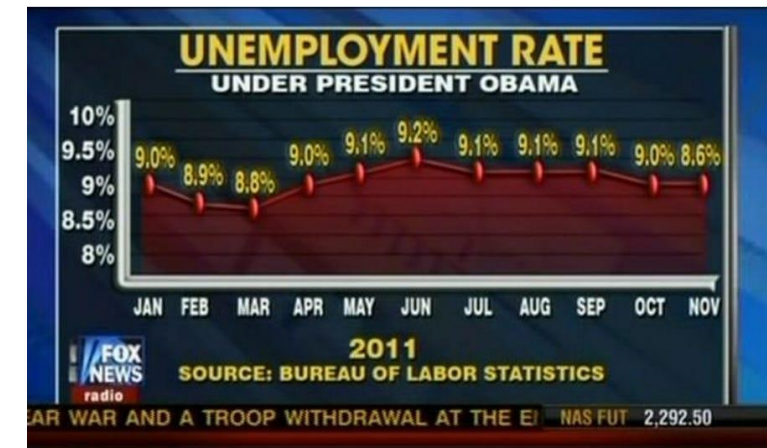
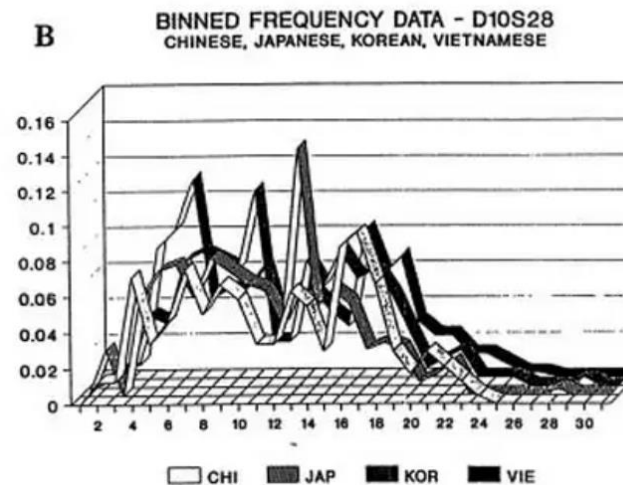
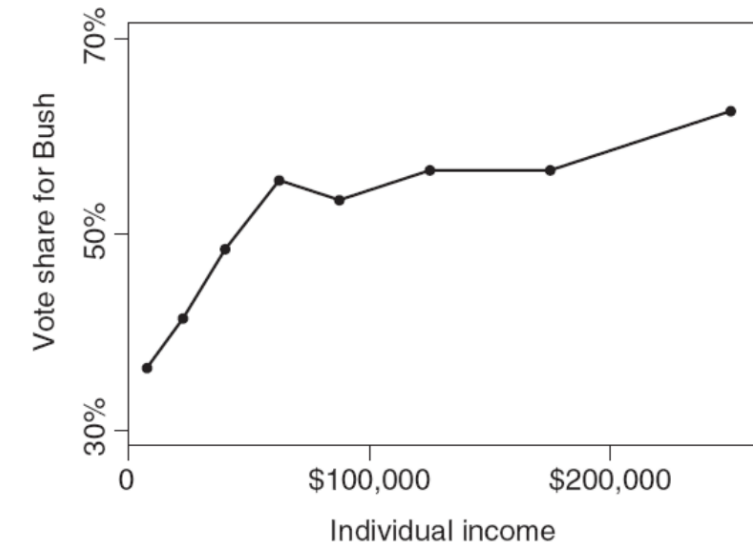
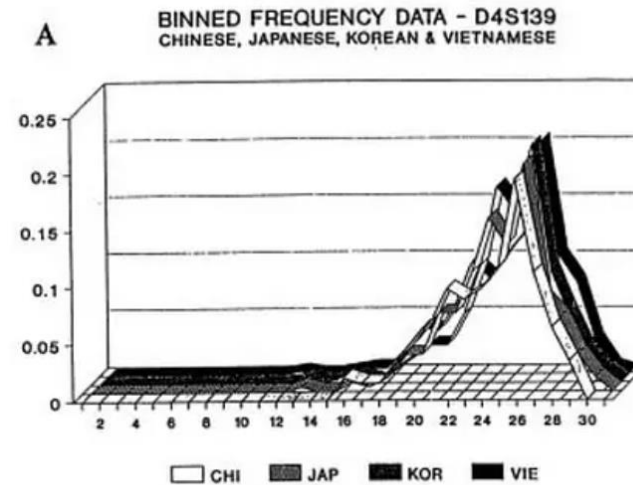
<https://www.dummies.com/education/math/statistics/how-histograms-can-misrepresent-statistical-data/>

GRAPH TYPES: LINE GRAPHS

What is wrong with this all these graphs?

Connect a line through the several points of an indicator

Useful if you need to visualise a trend across data



GRAPH TYPES: SCATTERPLOTS

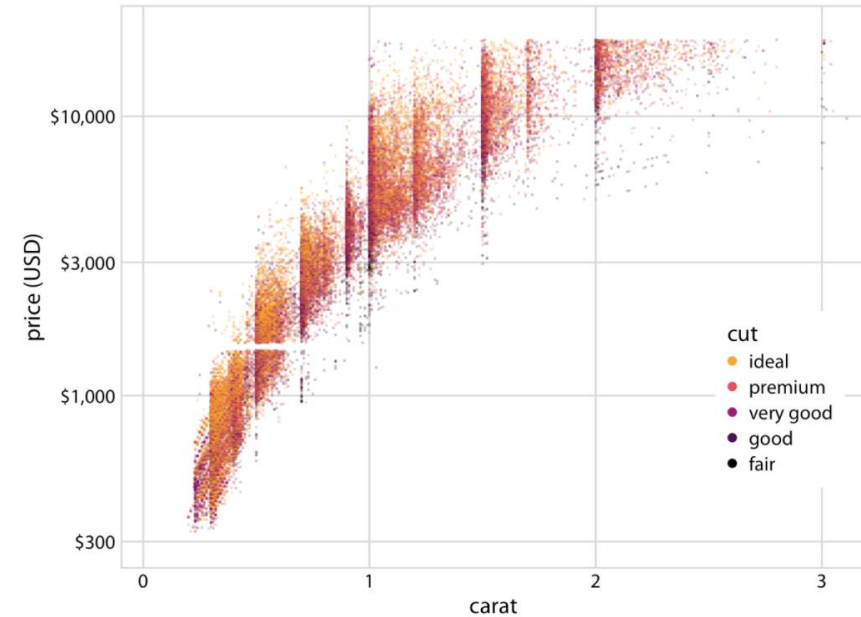
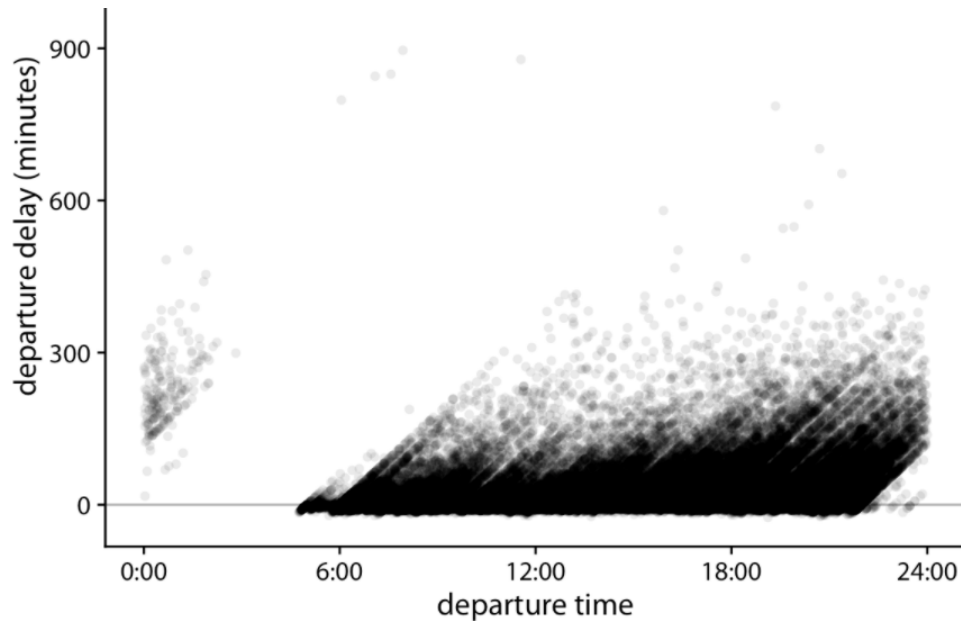
You can visualise the relationship between two interval variables that you are interested in

DV is plotted on the y-axis,

IV is depicted on the x-axis

You can identify outliers (=deviations)

What is wrong with these graphs?



Read more here: <https://clauswilke.com/dataviz/>

GRAPH TYPES: BOXPLOTS

Try to understand this one:

Display the distribution of different categories

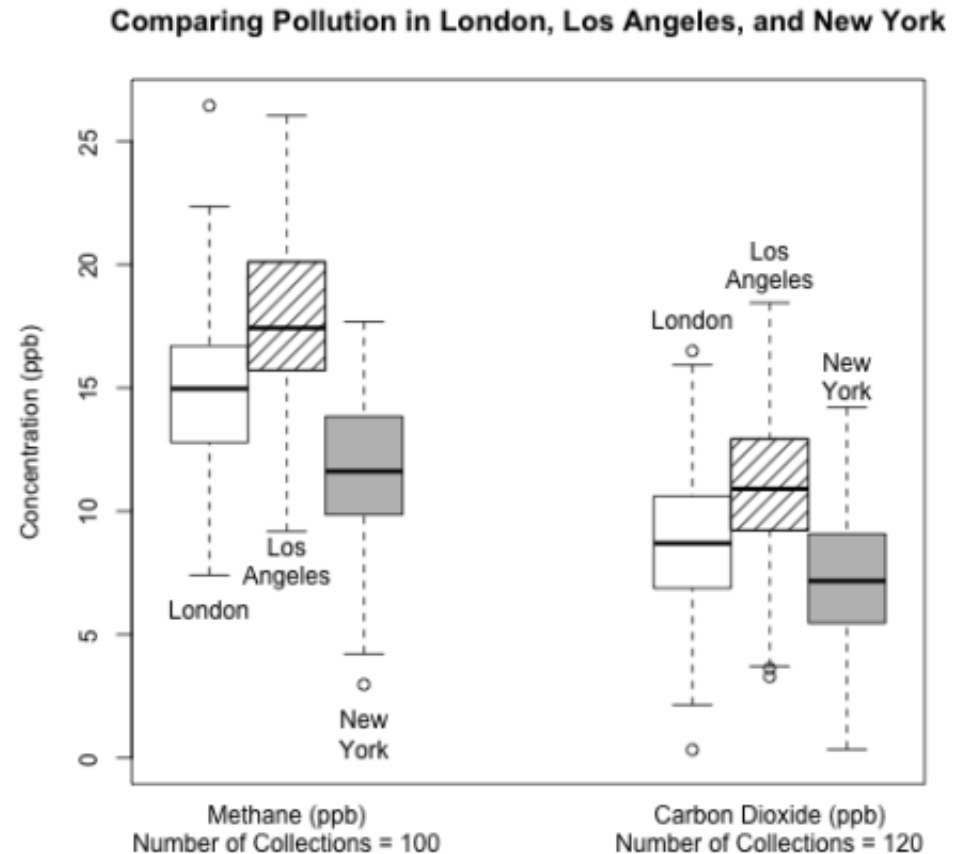
Visual summary of descriptive statistics

Centre line = median

Boxes = interquartile range (25th to 75th percentile)

The 'whiskers' extend to $\pm 1.5 \times \text{IQR}$

Points more extreme than these values (outliers) are represented with dot



Source: <https://www.r-bloggers.com/2014/04/side-by-side-box-plots-with-patterns-from-data-sets-stacked-by-reshape2-and-melt-in-r/>