Imperial College
Business School

Imperial means
Intelligent Business

Visions

– Often the simplest way to tell a story is a picture

by Ralf Martin (r.martin@imperial.ac.uk)

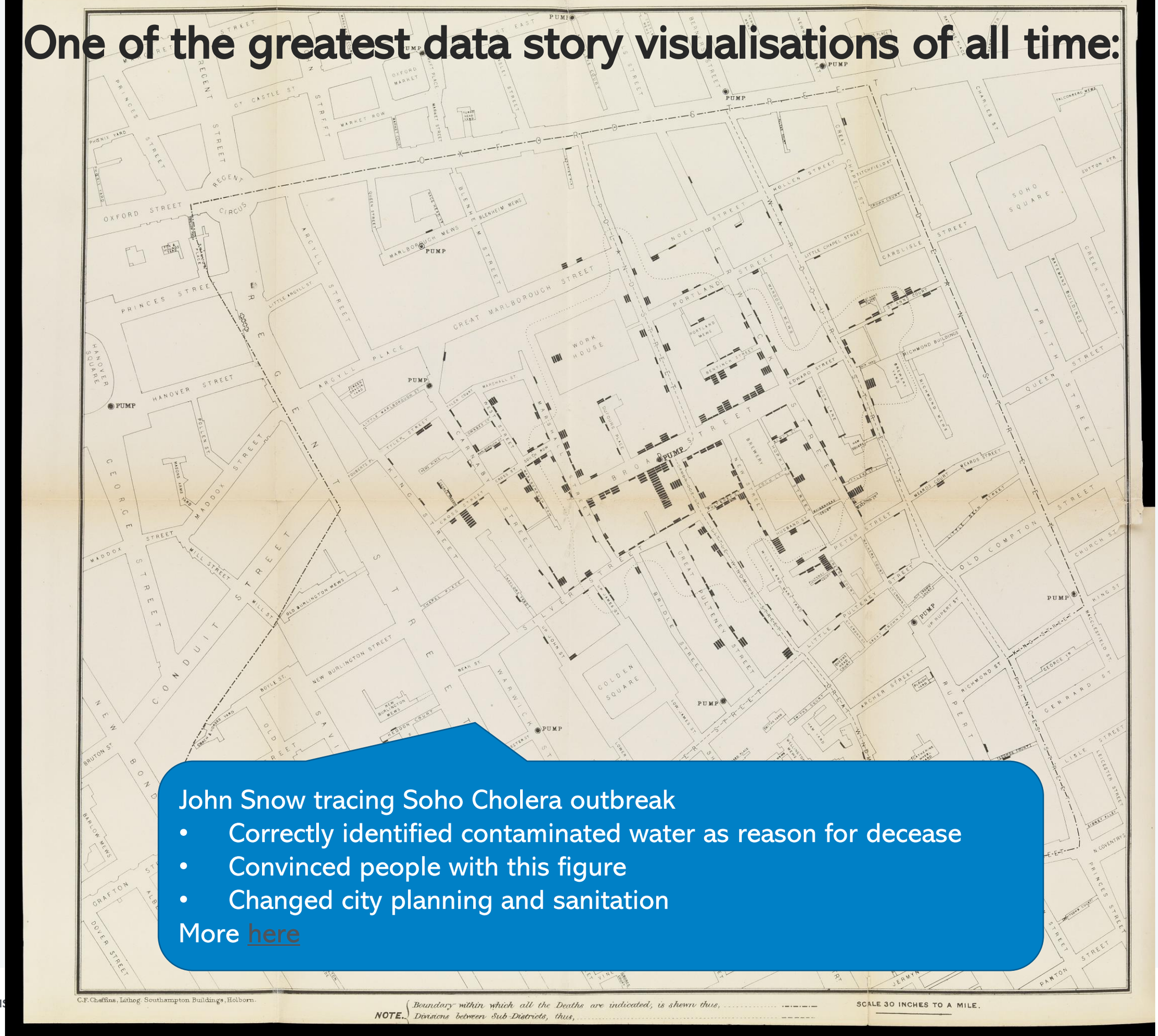
Plan for this lecture

The good, the bad, the ugly and how to make things better

- Good data visualisations are a great way of telling a story
- Let's look at some examples
- Let's learn some R commands for visualisations
- Let's try to make some new visualisations

In part based on lecture notes by Richard Davies

One of the greatest data story visualisations of all time:



John Snow tracing Soho Cholera outbreak

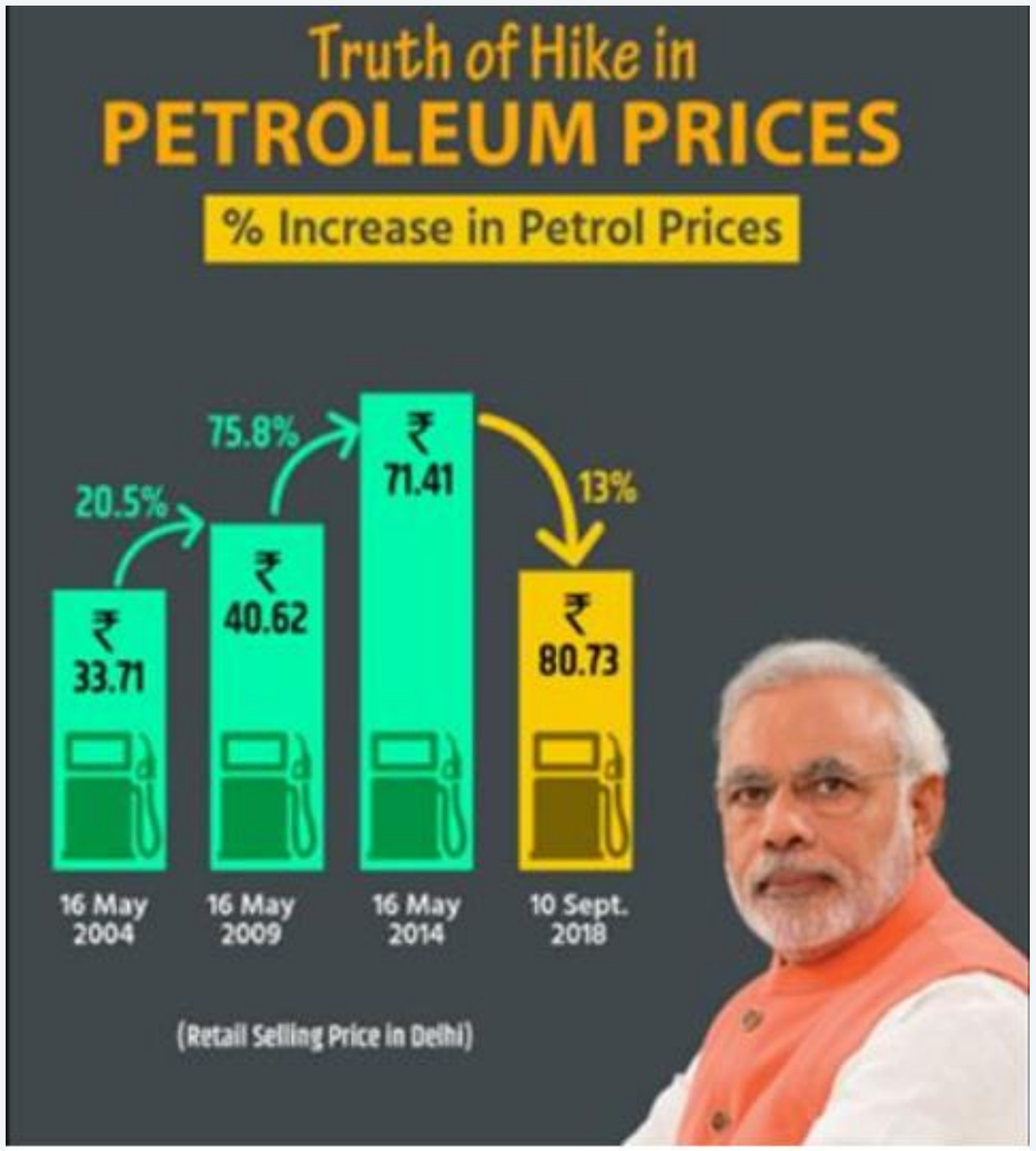
- Correctly identified contaminated water as reason for decease
- Convinced people with this figure
- Changed city planning and sanitation

More here

The bad - What can go wrong?

- Deliberate misleading

Pretending prices have come down when they haven't



More here

The bad - What can go wrong?

- Deliberate misleading
- Incompetence

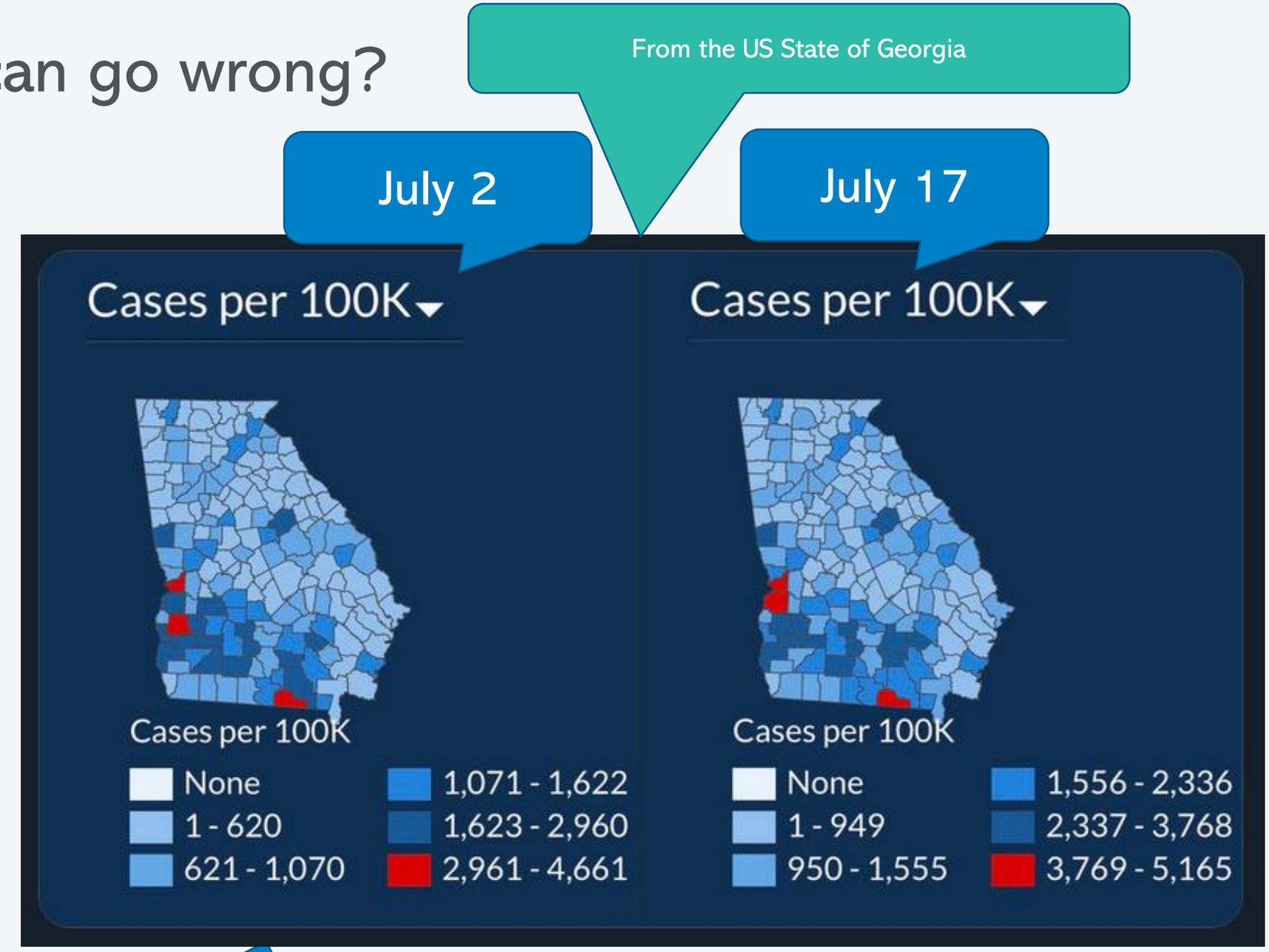
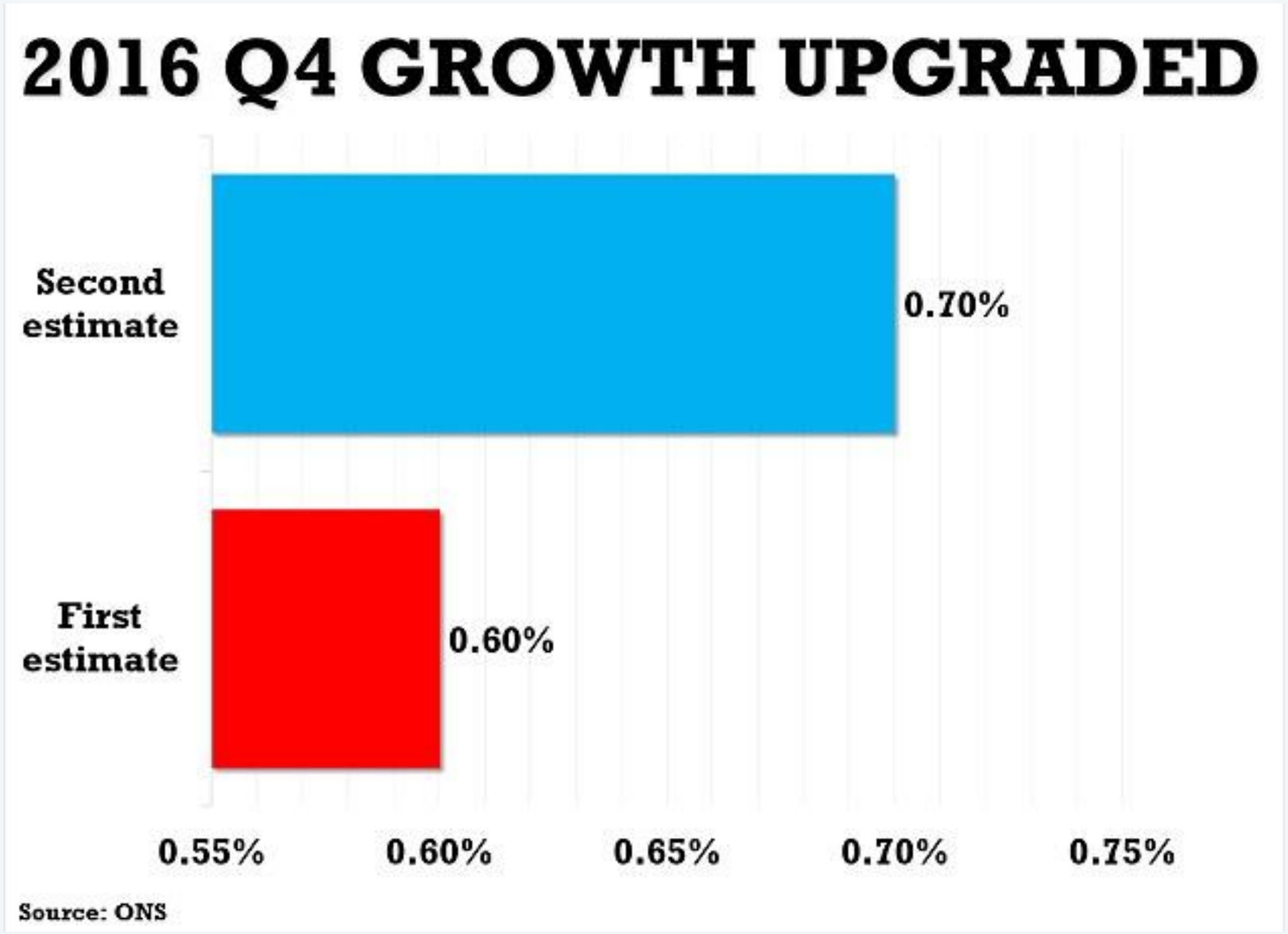


Figure seems to tell the story that COVID situation hasn't changed much (when it has)

More here

The bad - What can go wrong?

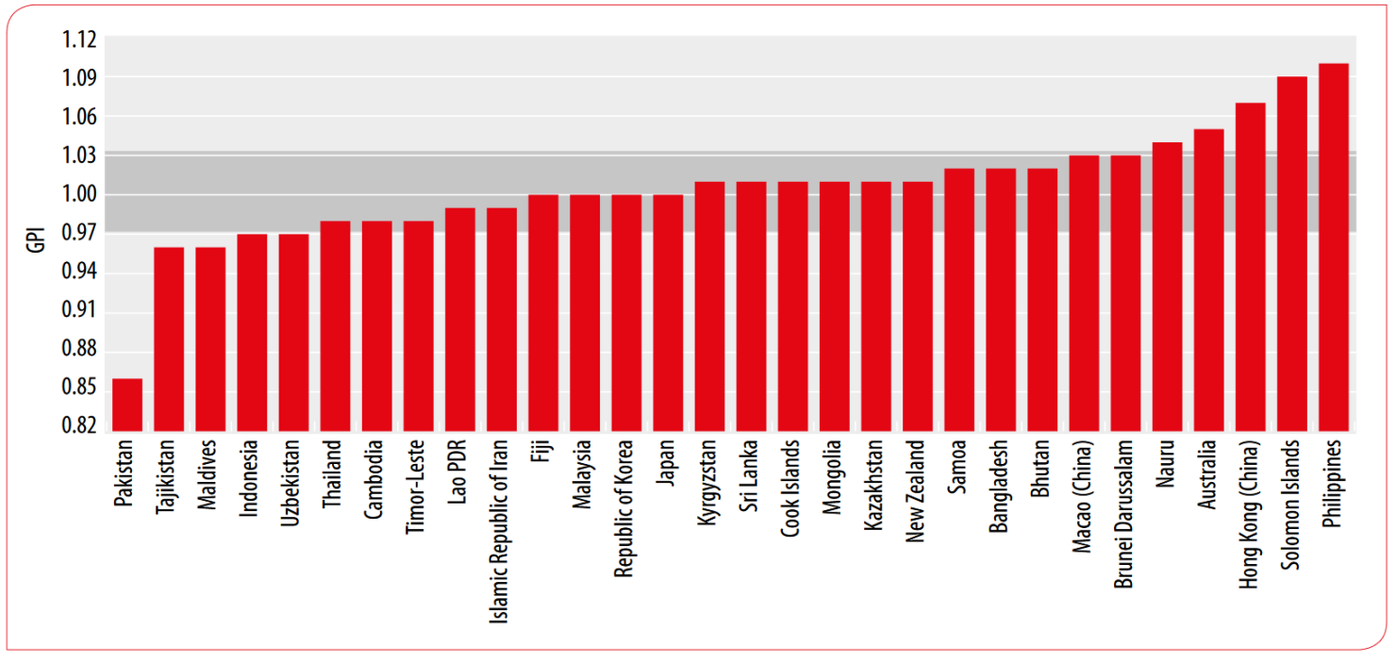
- Deliberately misleading



Pretending that something is a bigger deal (when it isn't)

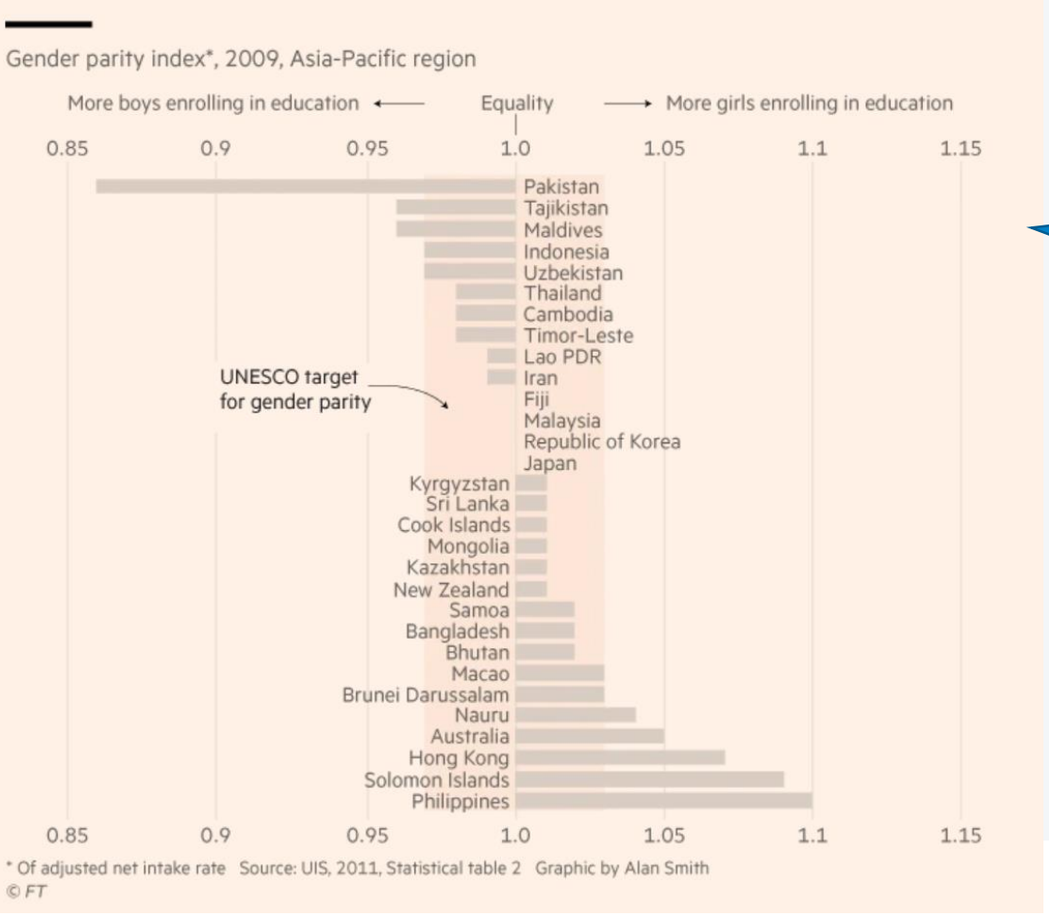
The bad – What can go wrong?

Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.

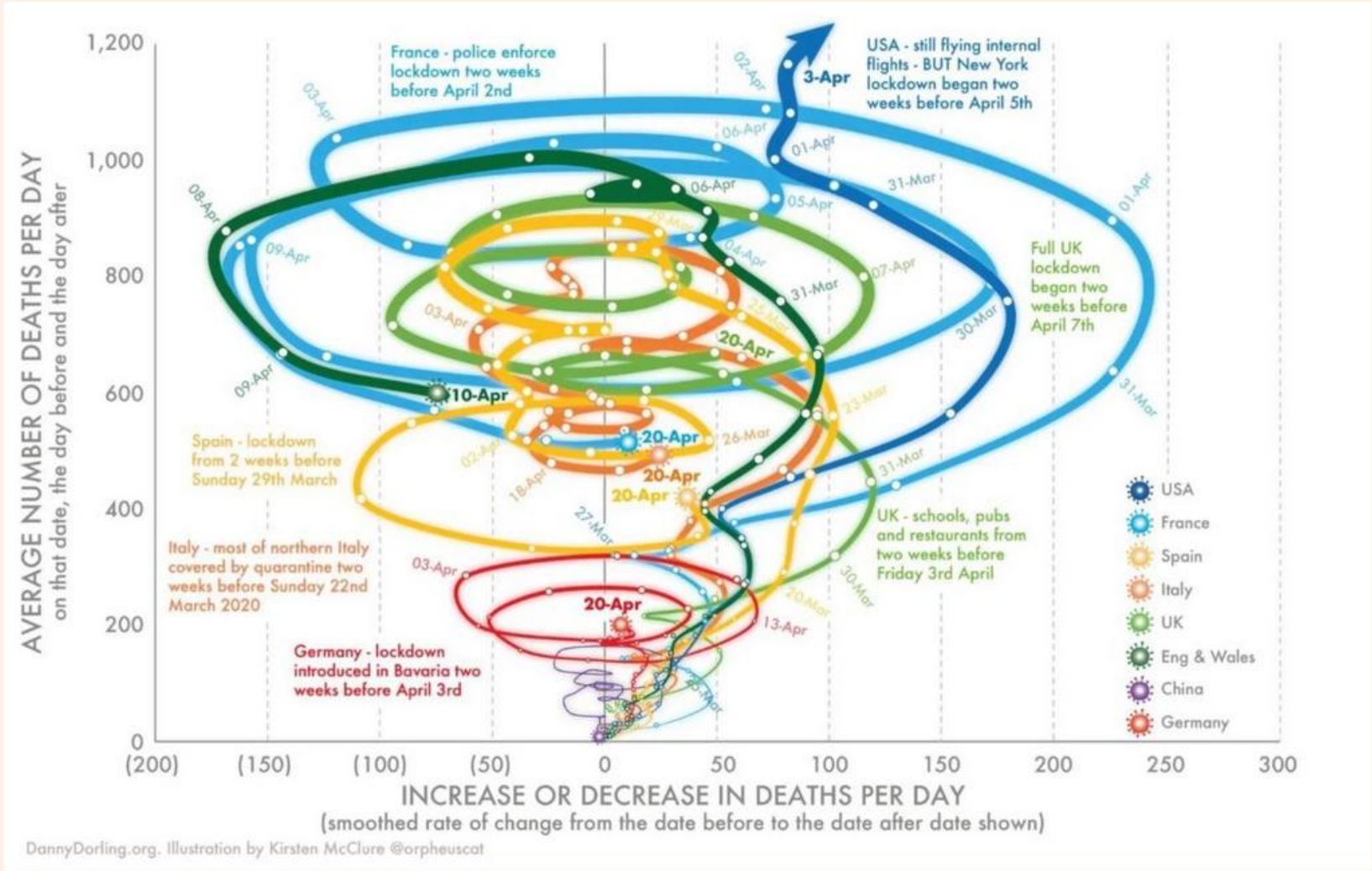
Looks like the Philippines are doing best?



This might be a better way (more discussion here)

The bad – What can go wrong?

There are axes of evil and then there's...

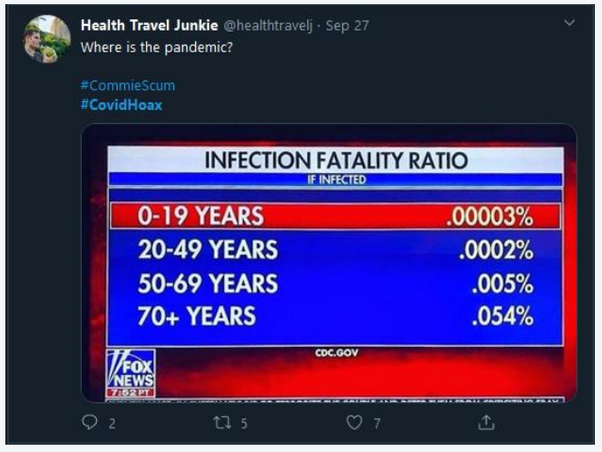


Looks kind of cool but what does it tell us?
For many more examples consult the FT's chart doctor

R visuals – Let's get some data for examples

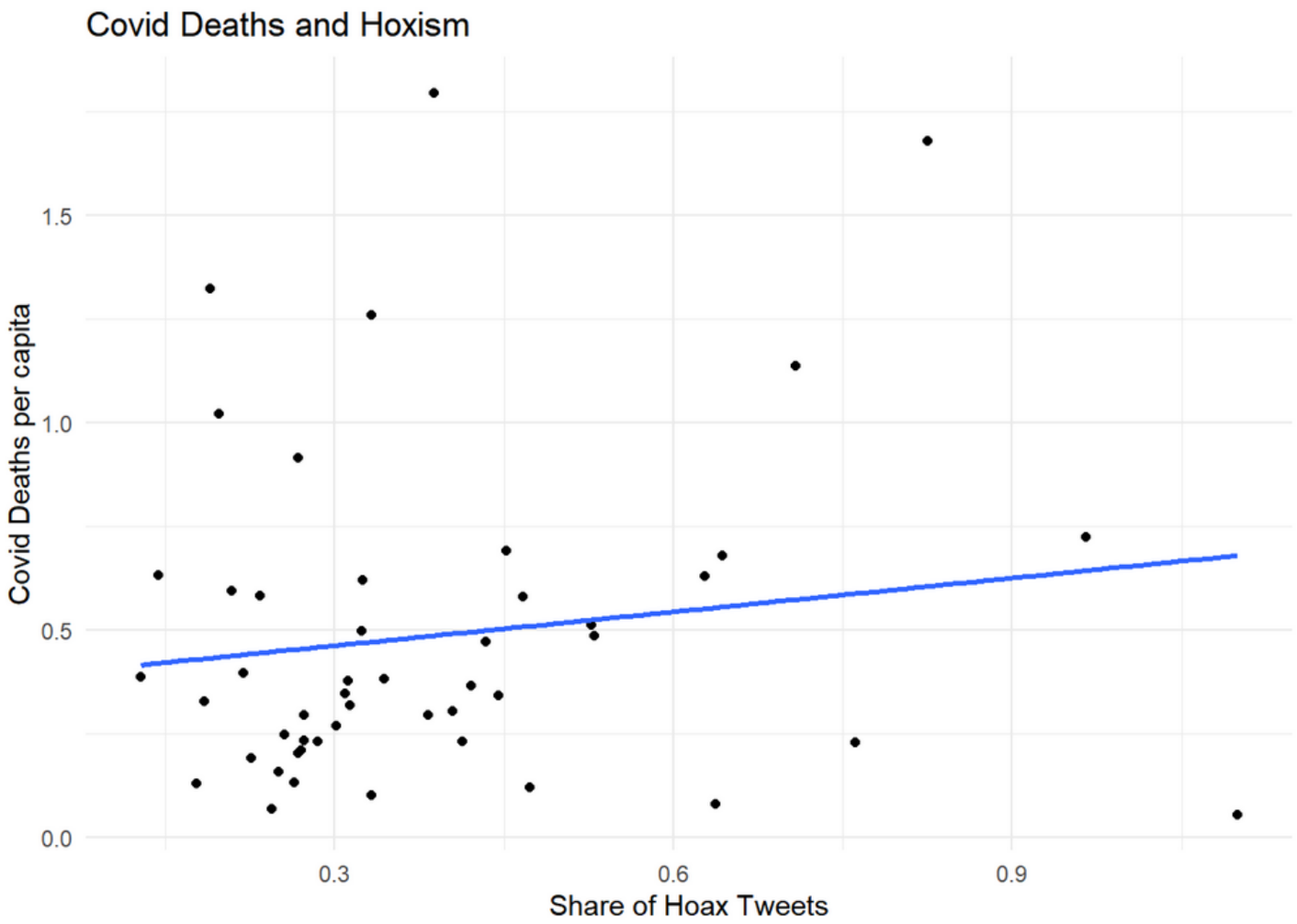
```
library(dplyr)
stats=read.csv("https://www.dropbox.com/s/8w4zbg40y84pnqk/statslong.csv?dl=1")

# Creating some extra variables
stats=stats%>%mutate(   pop=pop/1000,
                        hoaxshXdensity=(hoaxsh)*(density-mean(density)),
                        tweetsPCXdensity=(tweetsPC)*(density-mean(density))
                     )
```

- Data on COVID hoaxism
- Rmd file with code for this lecture
- html file



Scatter ggplot– The relationship between COVID hoaxism and deaths

```
library(ggplot2)
ggplot(stats,aes(x=hoaxsh, y=deathsPC))+
       geom_point() +
       theme_minimal()+
       xlab("Share of Hoax Tweets") +
       ylab("Covid Deaths per capita") +
       geom_smooth(method = "lm", se = FALSE)+
       ggtitle("Covid Deaths and Hoxism")
```



Adding twists to your scatter plot story

- If hoaxism causes deaths we might expect this to be worse in more densely populated regions

```
stats=stats %>%mutate(dens_quart=cut(density,
                                     breaks=quantile(density, probs=seq(0,1, by=0.25), na.rm=TRUE),
                                     include.lowest=TRUE))
```

Creating quartile bins of the population density variables

```
stats %>% group_by(dens_quart) %>% summarise(mean(deathsPC))
```

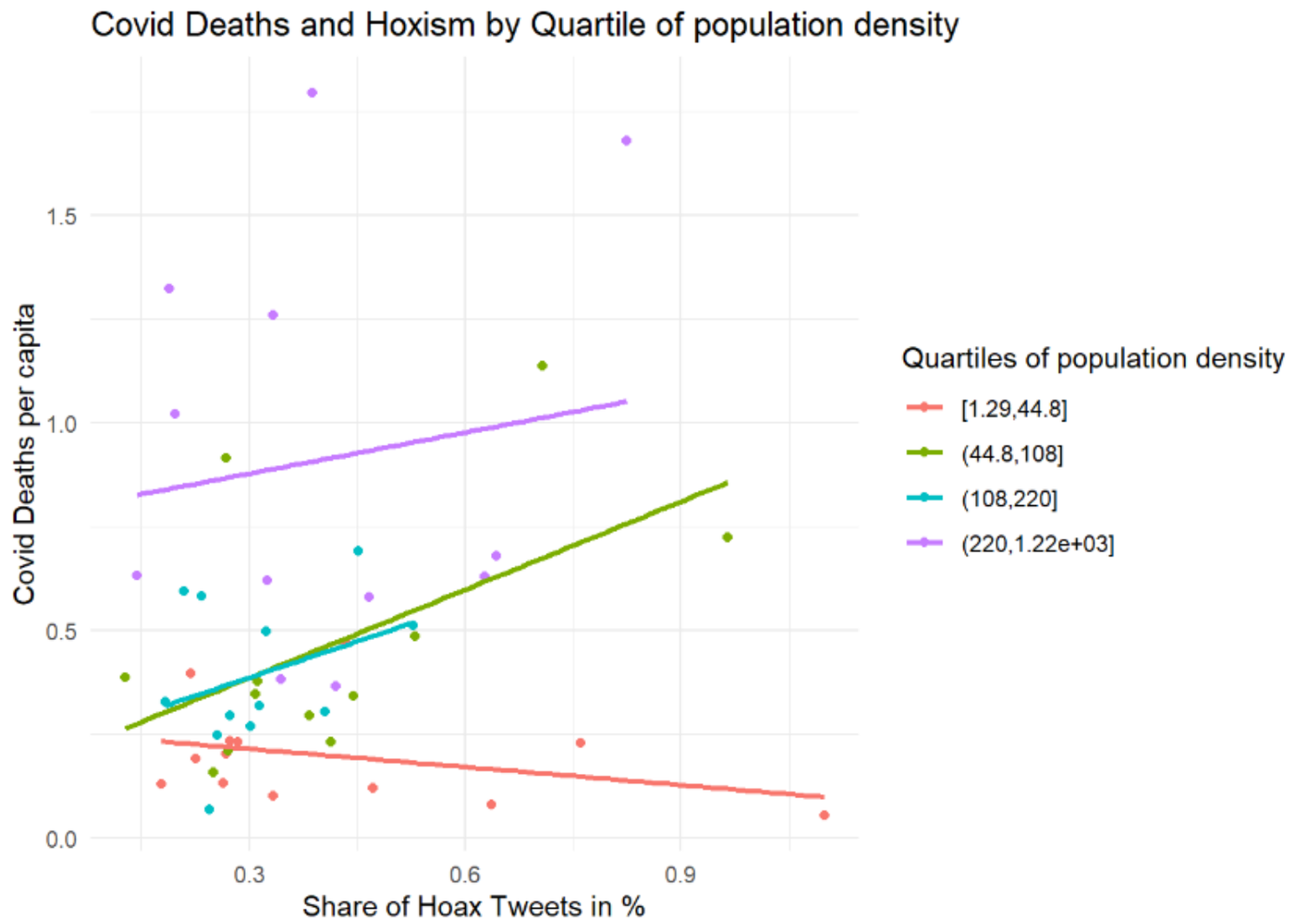
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 4 x 2
##   dens_quart     `mean(deathsPC)`
##   <fct>                     <dbl>
## 1 [1.29,44.8]               0.198
## 2 (44.8,108]                0.468
## 3 (108,220]                 0.393
## 4 (220,1.22e+03]            0.914
```

A lot more story with very little more code

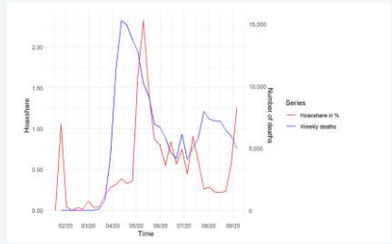
```
ggplot(stats,aes(x=hoaxsh, y=deathsPC, color=dens_quart))+
       geom_point() +
       theme_minimal()+
       xlab("Share of Hoax Tweets in %") +
       ylab("Covid Deaths per capita") +
       geom_smooth(method = "lm", se = FALSE)+
       ggtitle("Covid Deaths and Hoxism by Quartile of population density") +
       guides(color=guide_legend(title="Quartiles of population density"))
```

That's all

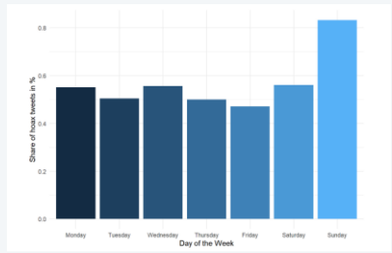


More visions (check here)

- Time Series
  - More is not always better

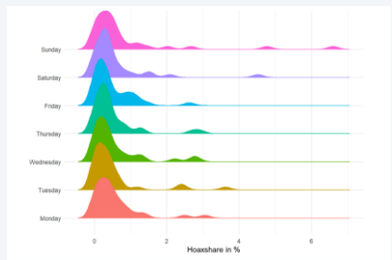


- Bar chart

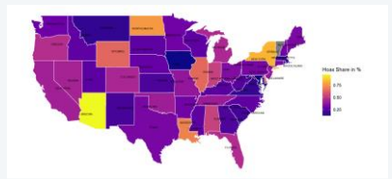


- Histogram
  - Density histogram

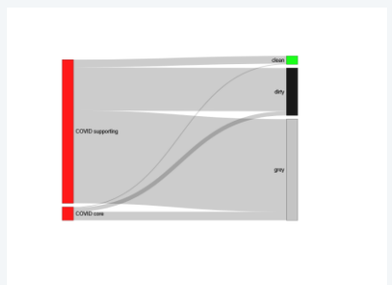
- Density Plot



- Map



- Integrating javascript



Takeaways

- R is great for doing visualisations

- Have a go yourself:
  - Find some data
  - Make a nice diagram with R Markdown
  - Tell some story with it (With R Markdown)
  - Post to RPubs as well as the Datathon Visualisation Challenge 2020

- To find data you can have a look at the Data Resources Channel

- If you are on twitter you can share @datastorieshub