

Why data visualization?

Abhijit Dasgupta

BIOF 339

The failed promise of data summaries

It is tempting to try and summarize data just by some data summaries, like means, medians or standard deviations.

Summaries, by their very nature, compress information

Which means **some information gets thrown out**

Visualization gives us a way to see patterns in the data that would not be obvious from data summaries

It also allows us to use our natural visual ability of pattern recognition to understand our data

The failed promise of data summaries

Anscombe's data

```
The downloaded binary packages are in  
/var/folders/k4/xvcmx4yx76xdbl41zy3hq8rc0000gn/T/
```

Statistic	Value
mean(x)	9
mean(y)	7.5
var(x)	11
var(y)	4.13
cor(x,y)	0.82

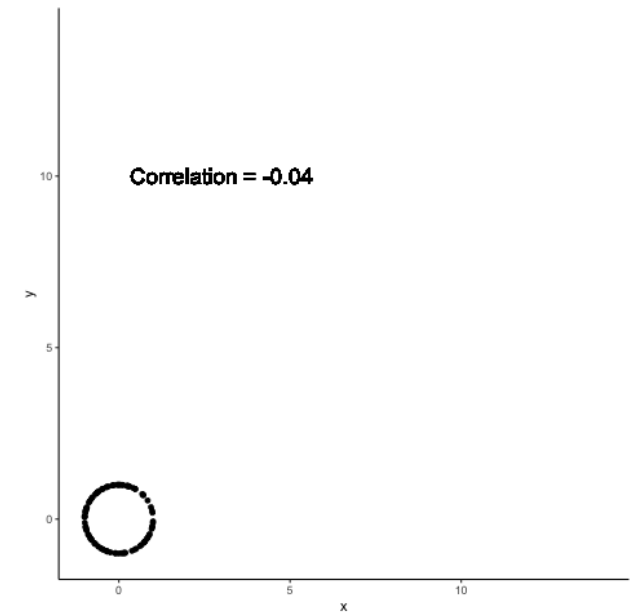
The variables for each data set have the same values of data summaries

The failed promise of data summaries

The **DataSaurus** dozen

The failed promise of data summaries

A single point can completely change the computed correlation



The failed promise of data summaries

Data summaries are meant to help distinguish between different data sets

Both Anscombe and Datasaurus show that this promise is not met by standard data summaries

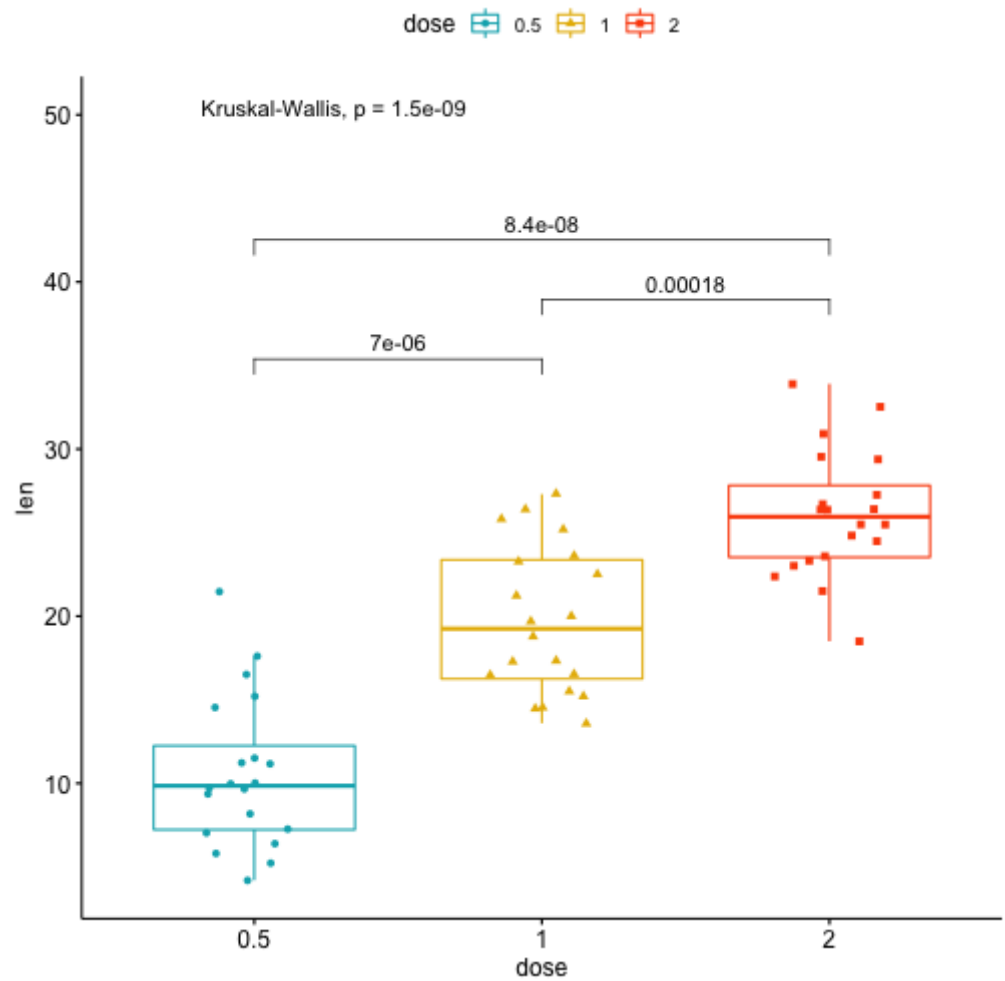
The previous example shows how a single point can change data summaries

Why visualize data?

- Summary statistics cannot always distinguish datasets
- Take advantage of humans' ability to visually recognize and remember patterns
- Find discrepancies in the data more easily

Some examples

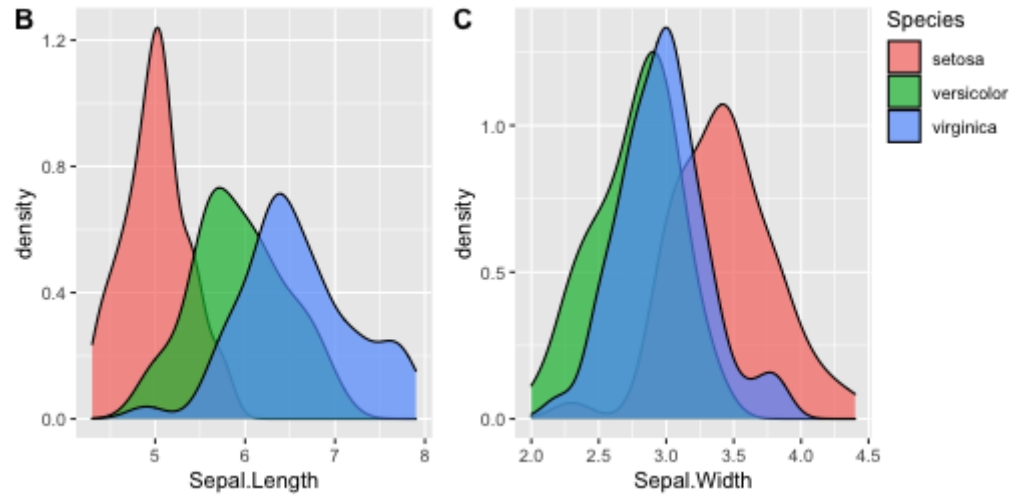
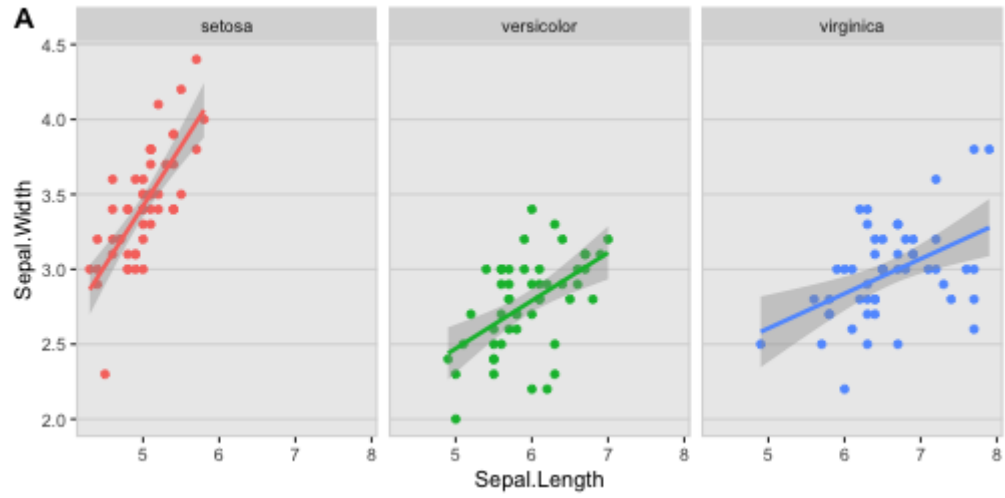
Gallery



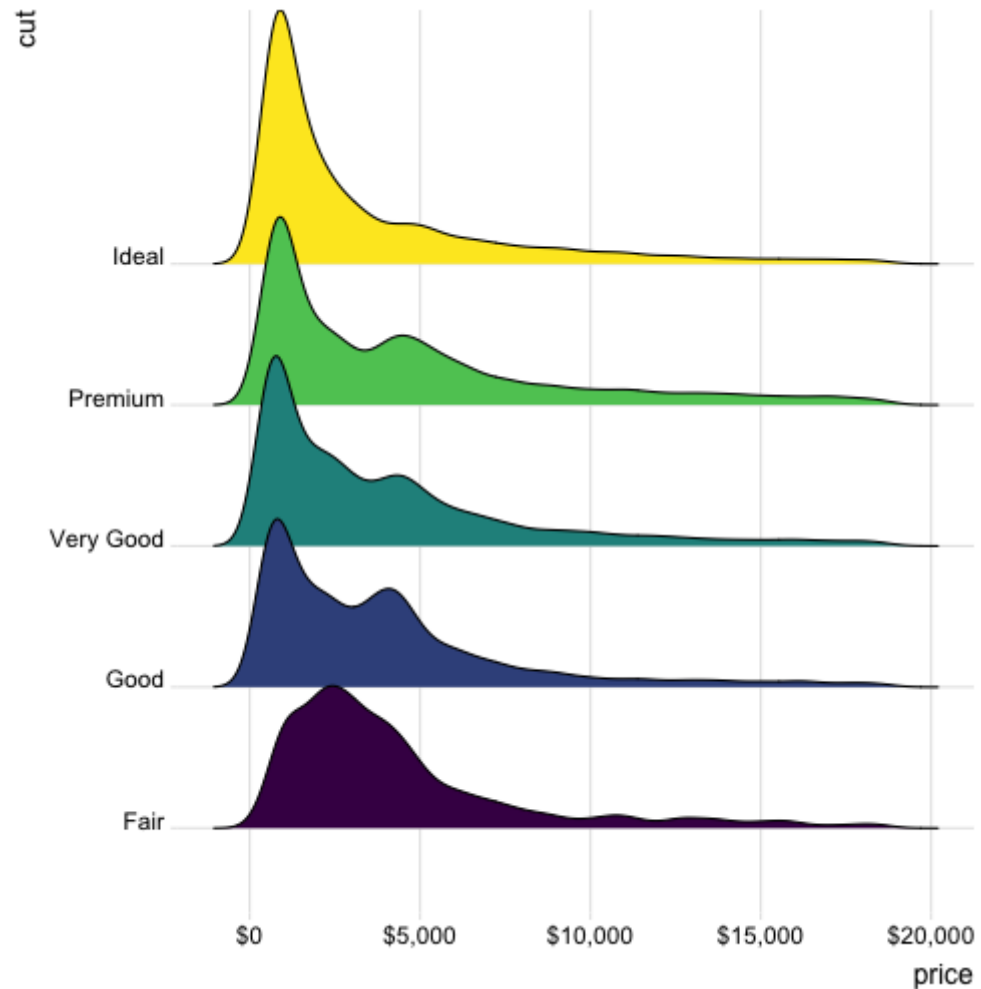
Gallery

Kaplan-Meier plots

Gallery



Gallery



Gallery

Manhattan plot

The downloaded binary packages are in
`/var/folders/k4/xvcmx4yx76xdb141zy3hq8rc000gn/T//Rtmpnj8c9g/downloaded_packages`

Gallery

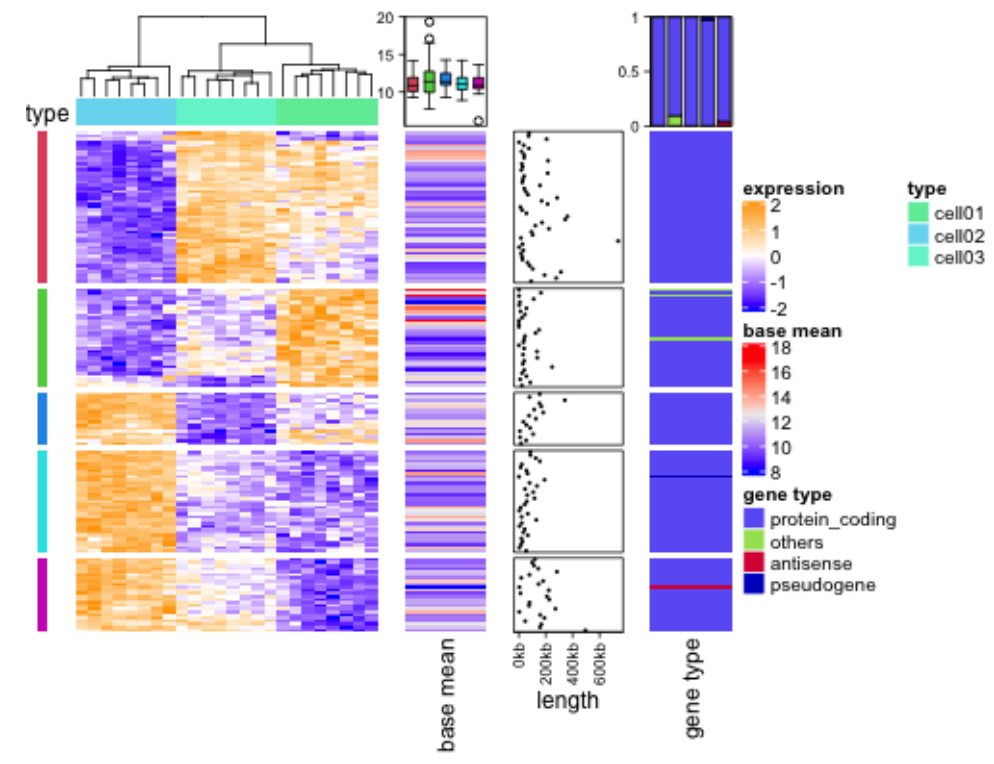
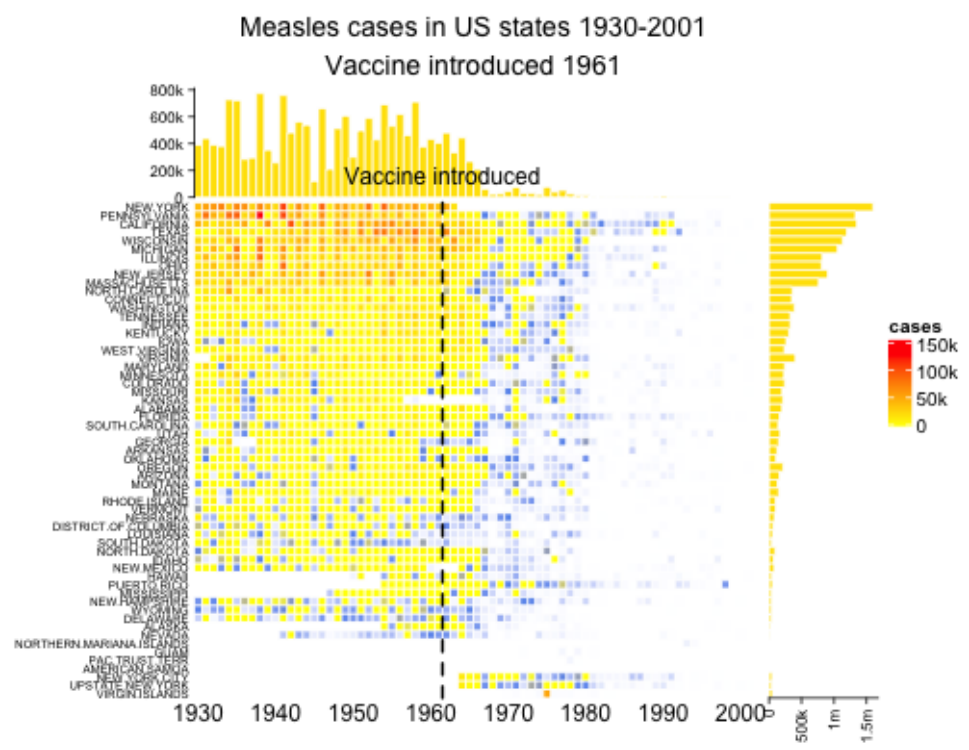
Circular Manhattan plot

Gallery

Maps

Gallery

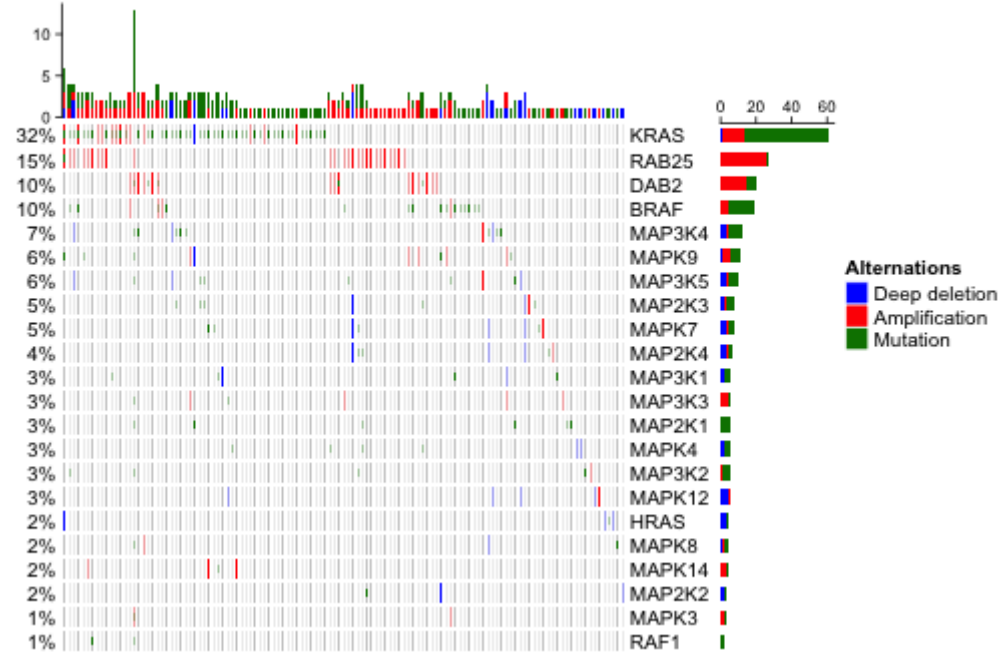
Heatmap



Gallery

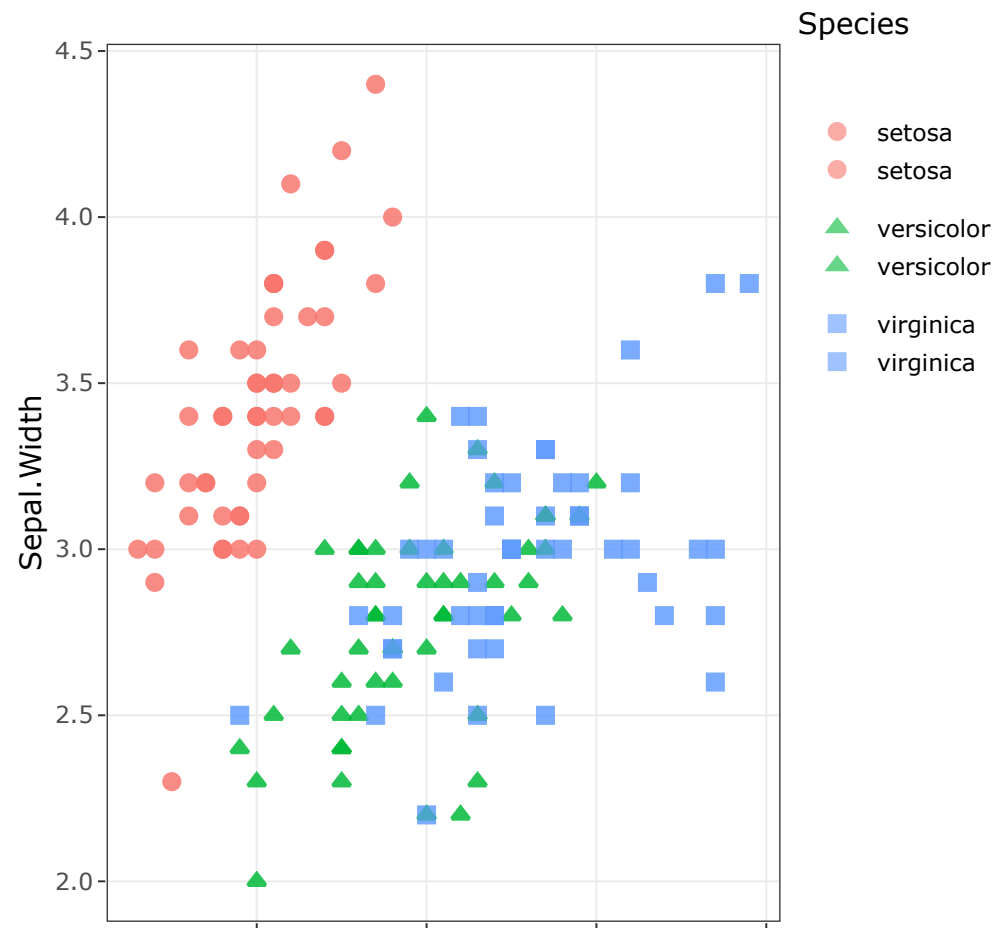
OncoPrint

TCGA Lung Adenocarcinoma, genes in Ras Raf MEK JNK signalling



Gallery

Interactive graphs

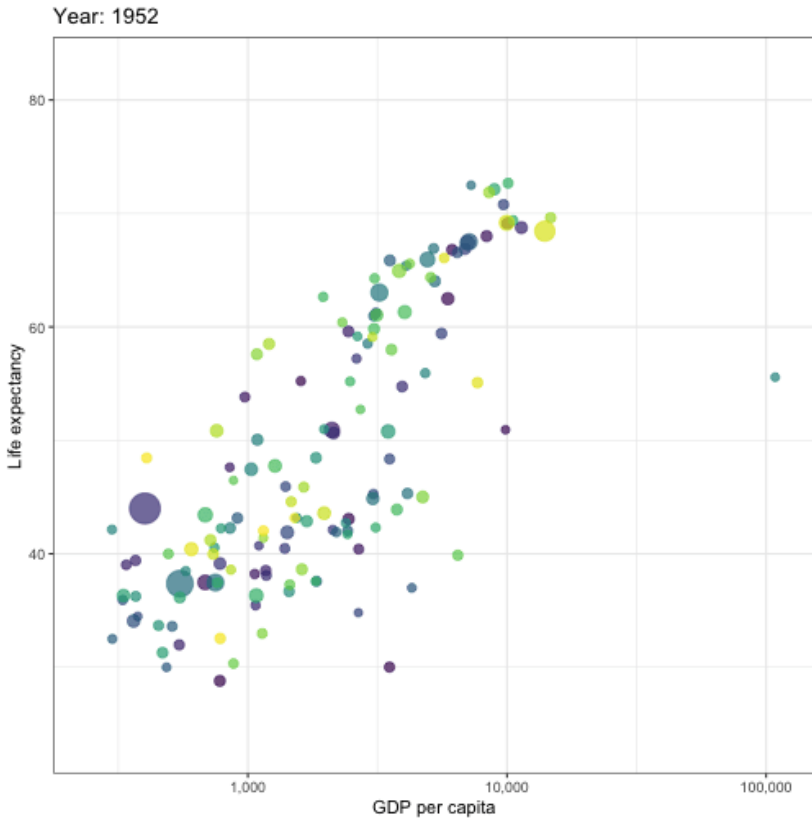


Gallery

Network graphs

Gallery

Animated graphs



**For more in-depth looks at data viz,
consider BIOF 439 in the Spring**