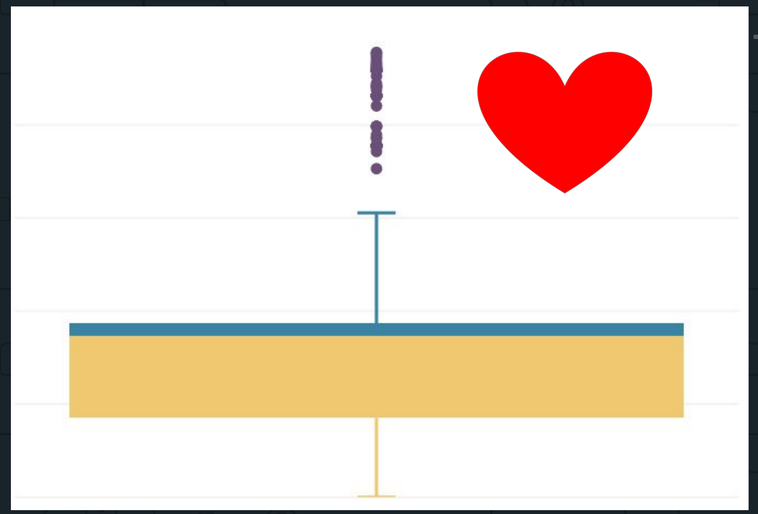
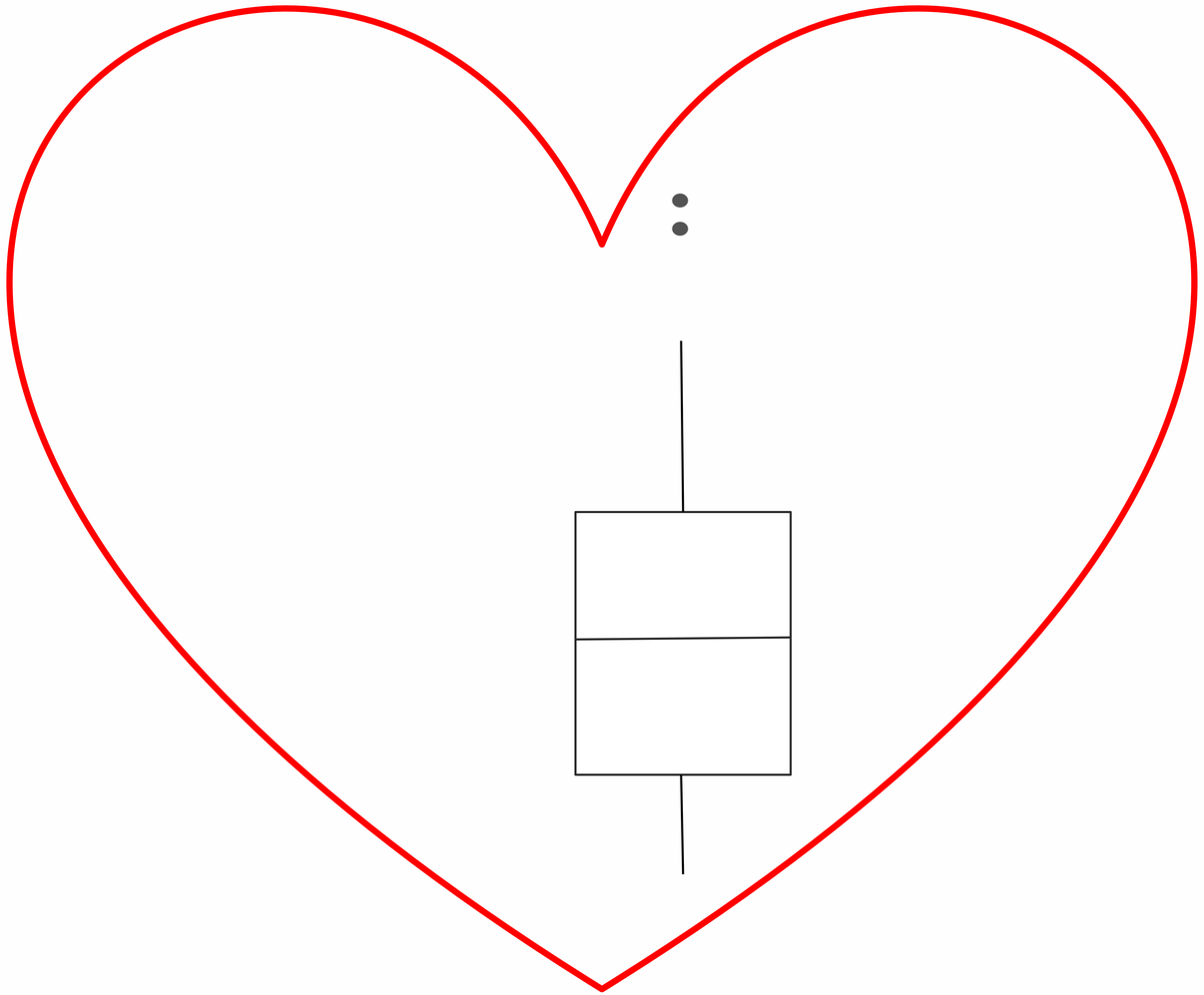
Pingboard

A Love Letter to the Boxplot

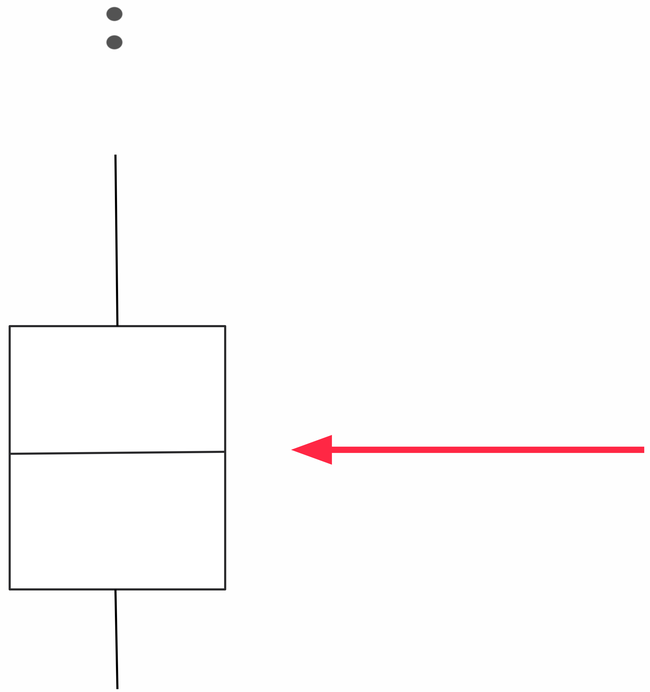


Outline

- Introduction to the Boxplot
- Drawing Boxplots with Computers
- Example Boxplots

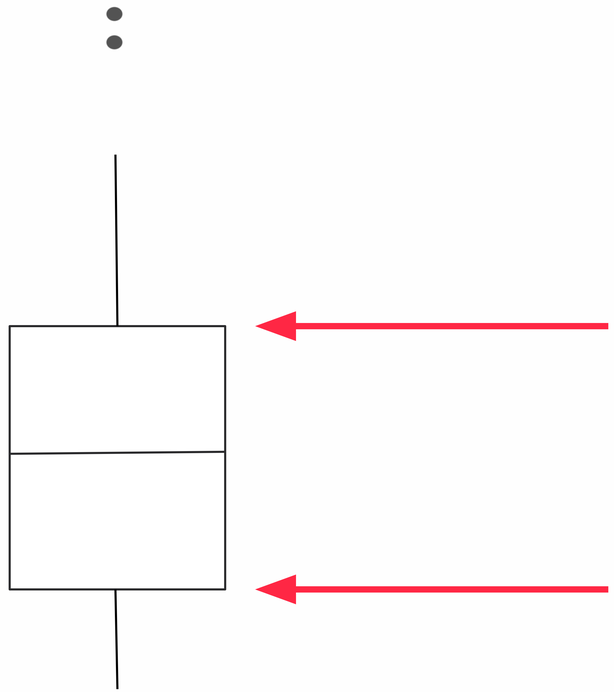


Anatomy of a Boxplot



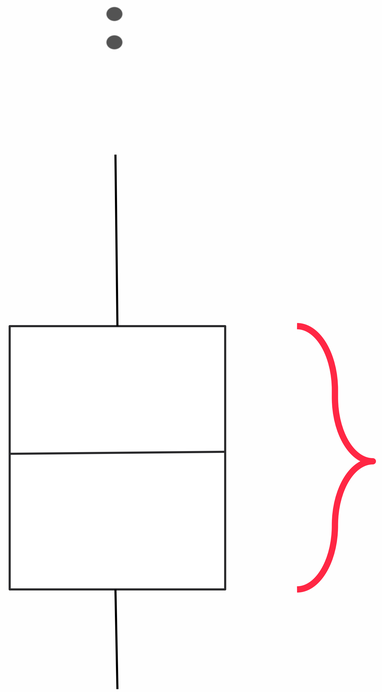
You've probably met the **median**, the halfway point of the distribution. Half above, half below

Anatomy of a Boxplot



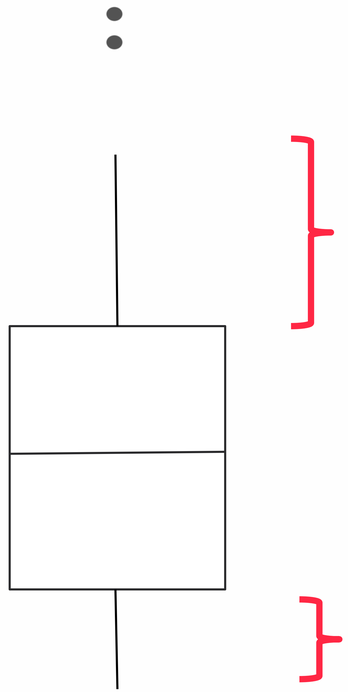
The **Upper** and **Lower Quartiles** are like the median - the divide their two halves into halves. (They are sometime known as the **Fourths** or the **Hinges**.)

Anatomy of a Boxplot



The area between the Upper and Lower Quartile is the **Interquartile Range** and it's a perfectly good measure of the spread of a distribution - it contains the middle half.

Anatomy of a Boxplot

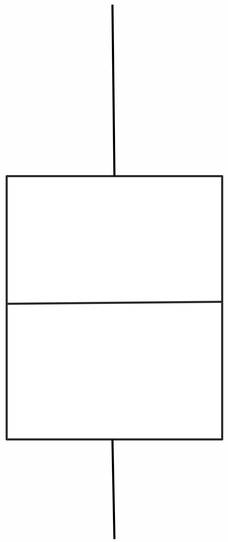


The **Whiskers** of the plot extend to the furthest point that is within 1.5*Interquartile Range past the end of the box

Anatomy of a Boxplot

: }

Anything past the whiskers is called an **Outlier** and gets its own mark on the graph

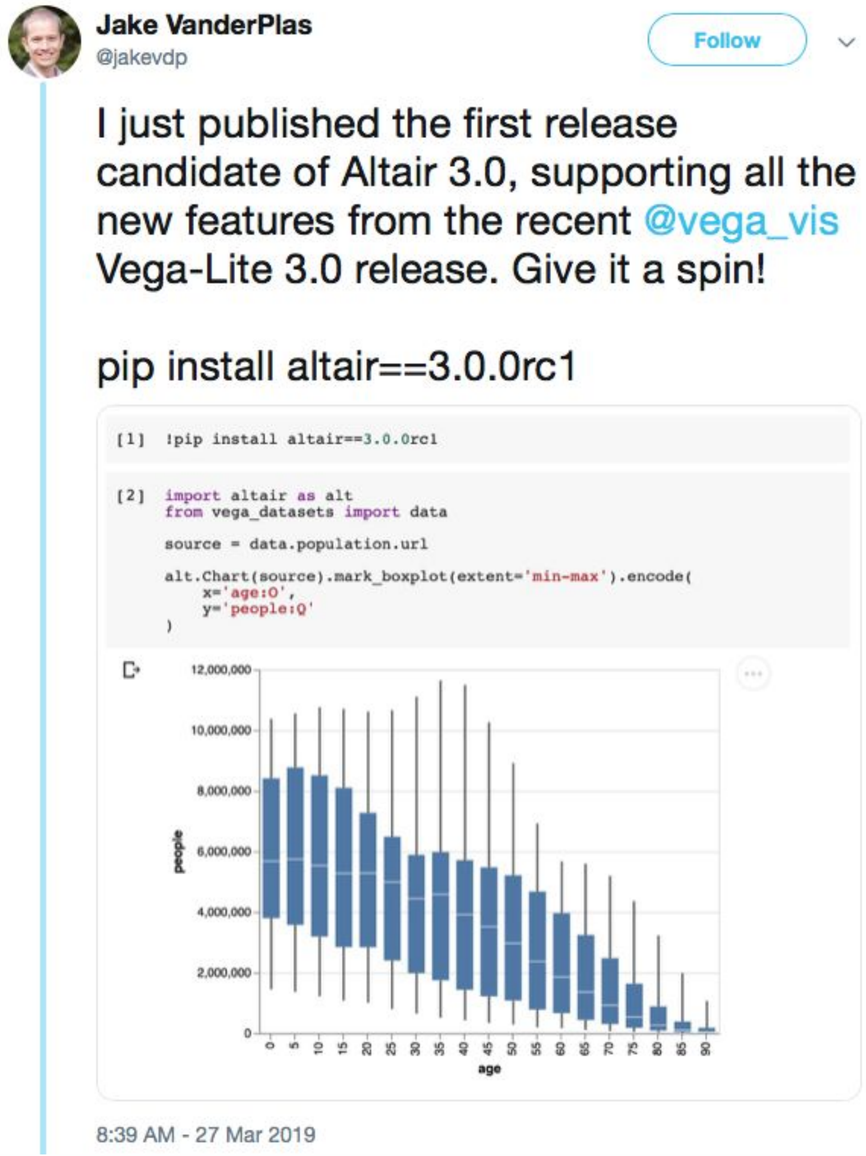


Why Boxplots? 

- Because people want to know what typically happens, and giving them the average is just rudely inadequate
- Easily seeing the range of the data - and finding unexpected outliers
- Quickly comparing groups

Drawing Boxplots with Computers

Lots of Tools to Create Boxplots



Charting Tools

Tool	Supports Boxplots
Tableau	Yes 
Chartio	Yes 
Power BI	Can be Imported
Minitab	Yes

https://www.kaggle.com/abcsds/pokemon

via

https://toolbox.google.com/datasetsearch

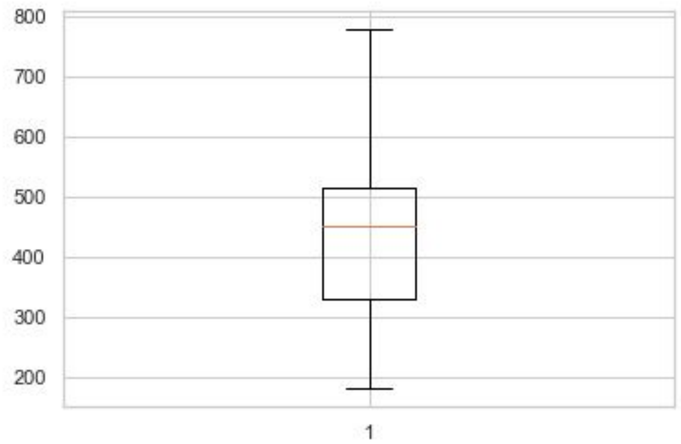
Python: Matplotlib and Pandas

https://matplotlib.org/api/_as_gen/matplotlib.pyplot.boxplot.html

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.boxplot.html

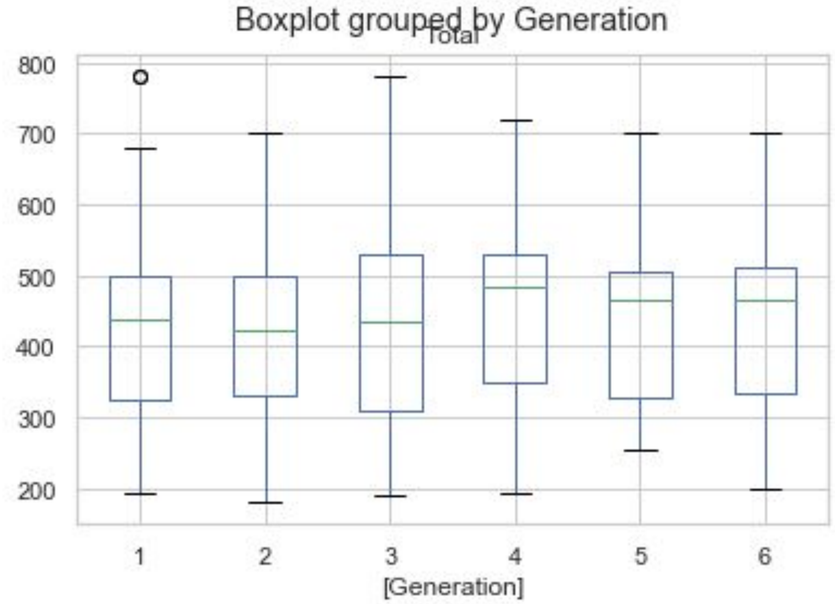
```
fig1, ax1 = plt.subplots()
ax1.boxplot(pokemon['Total'])
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x101626710>,
  <matplotlib.lines.Line2D at 0x101626a90>],
 'caps': [<matplotlib.lines.Line2D at 0x101626e10>,
  <matplotlib.lines.Line2D at 0x101626f28>],
 'boxes': [<matplotlib.lines.Line2D at 0x101626320>],
 'medians': [<matplotlib.lines.Line2D at 0x101628550>],
 'fliers': [<matplotlib.lines.Line2D at 0x1016288d0>],
 'means': []}
```




```
pokemon[['Total', 'Generation']].boxplot(by='Generation')
```

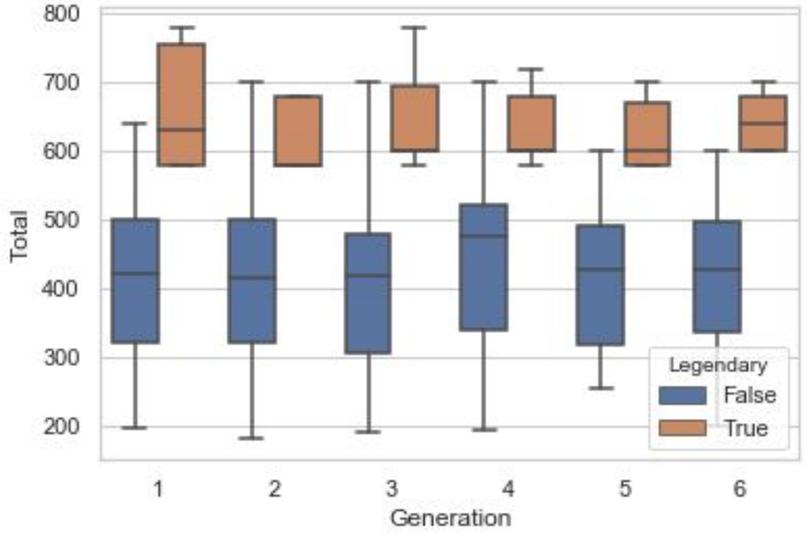
```
<matplotlib.axes._subplots.AxesSubplot at 0x110b64048>
```



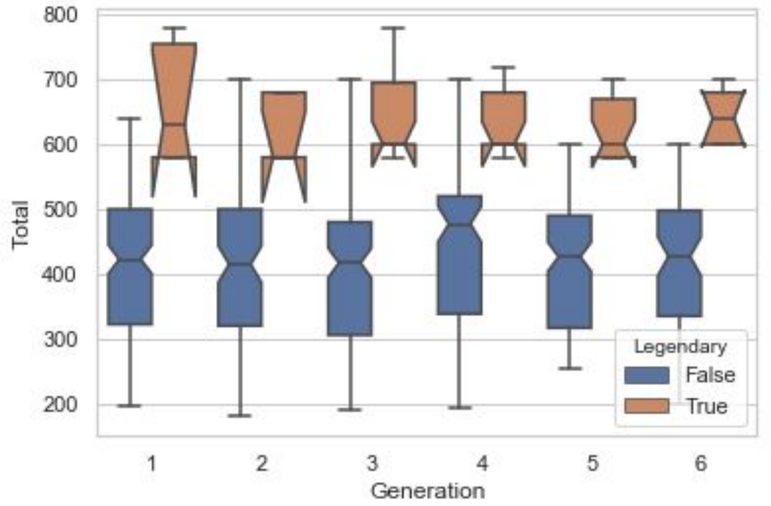
Python: Seaborn

https://seaborn.pydata.org/generated/seaborn.boxplot.html

```
ax = sns.boxplot(y='Total', x='Generation', hue='Legendary', data=pokemon)
```



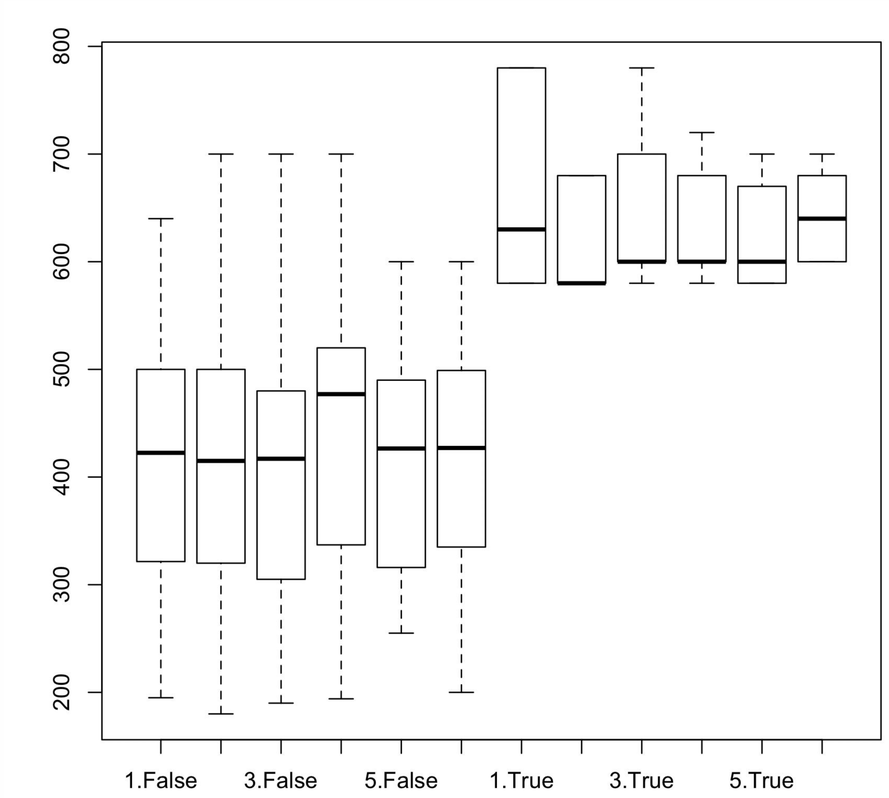

```
ax = sns.boxplot(y='Total', x='Generation', hue='Legendary', notch=True, data=pokemon)
```



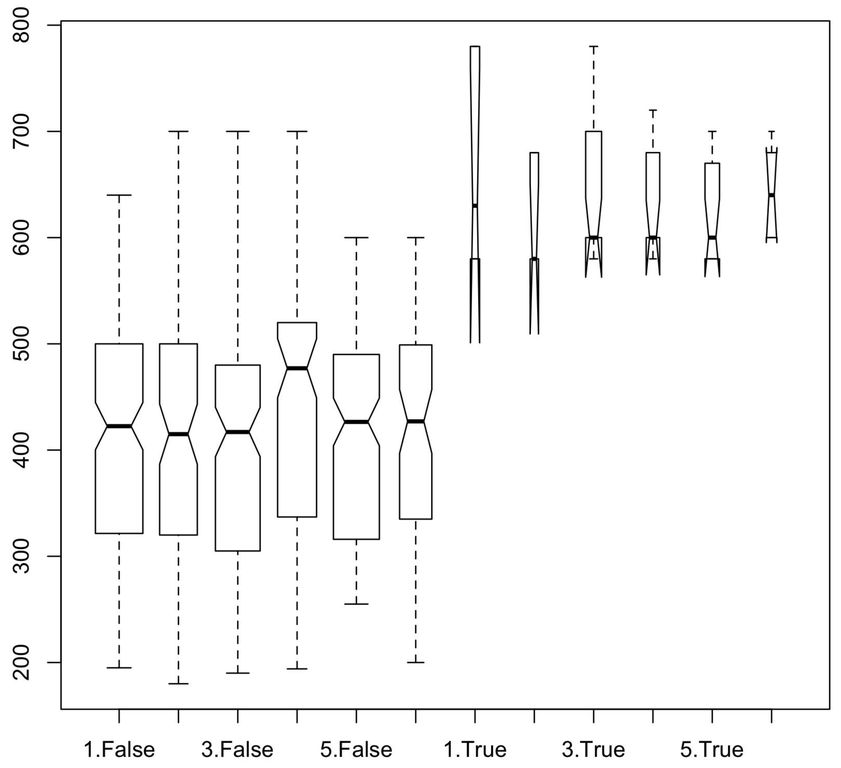
R: base graphics

https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/boxplot.html

```
boxplot(Total~Generation+Legendary, data=pokemon)
```

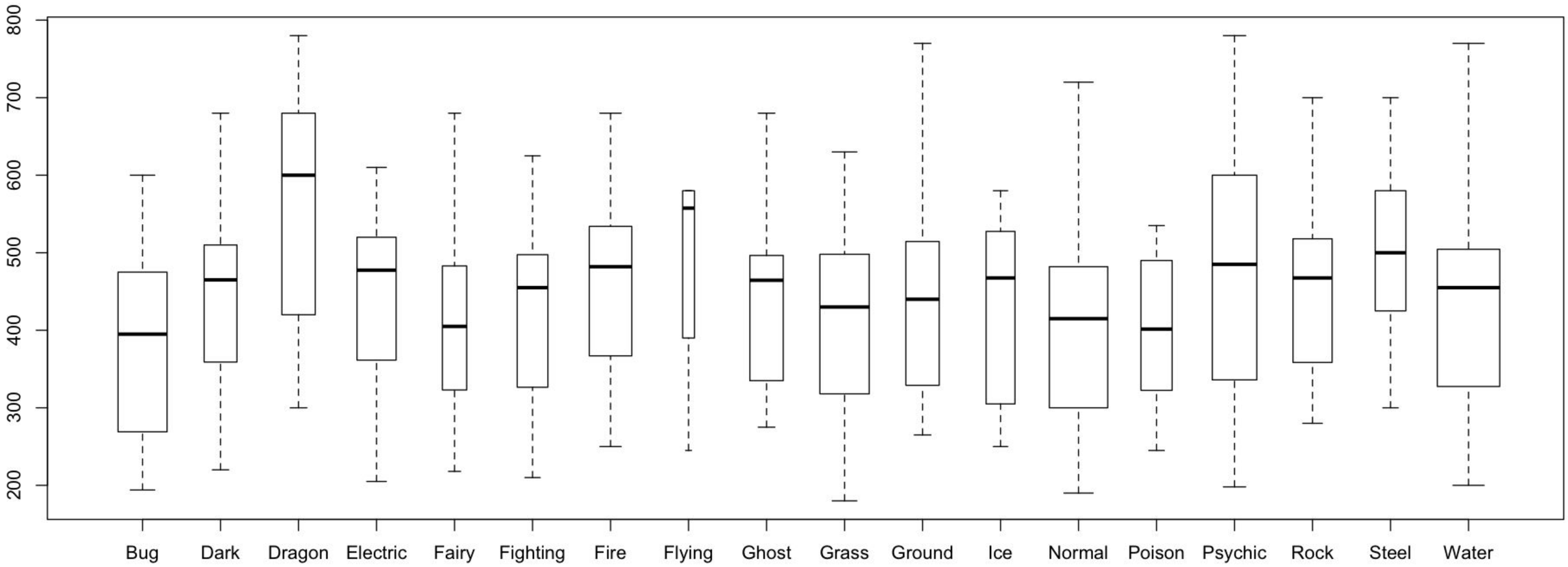



```
boxplot(Total~Generation+Legendary, data=pokemon,
varwidth=TRUE, notch=TRUE)
```



R: base graphics

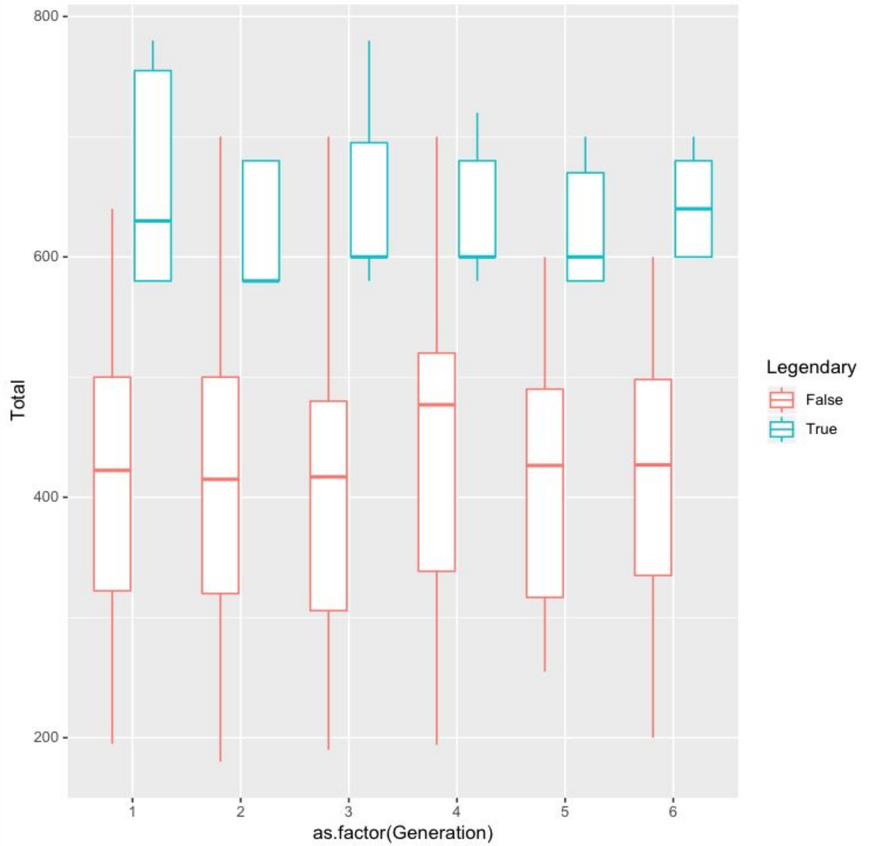
```
boxplot(Total~Type.1, data=pokemon, varwidth=TRUE)
```



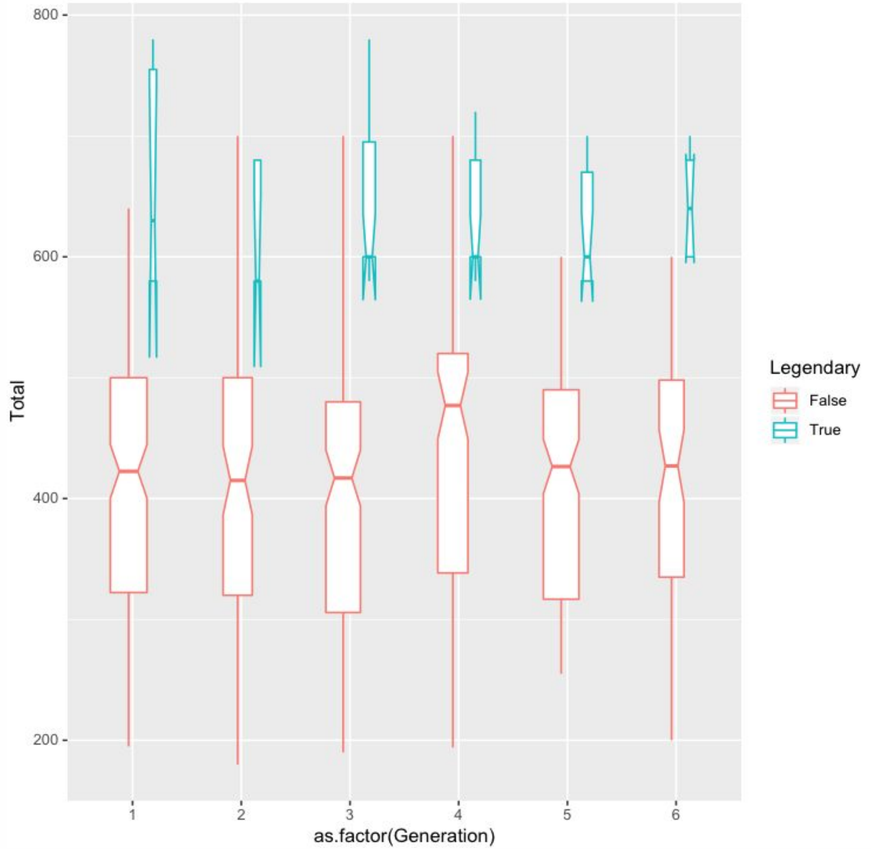
R: ggplot2

https://ggplot2.tidyverse.org/reference/geom_boxplot.html

```
p <- ggplot(pokemon, aes(x=as.factor(Generation), y=Total, color=Legendary)) + geom_boxplot()
```



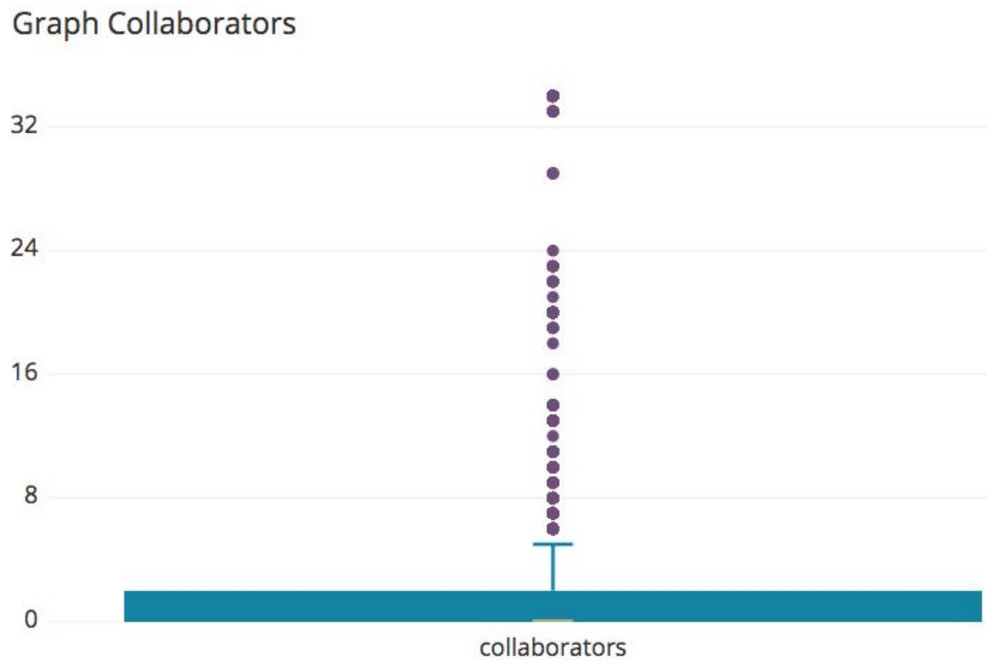

```
p <- ggplot(pokemon, aes(x=as.factor(Generation), y=Total, color=Legendary)) + geom_boxplot(notch=TRUE, varwidth=TRUE)
```

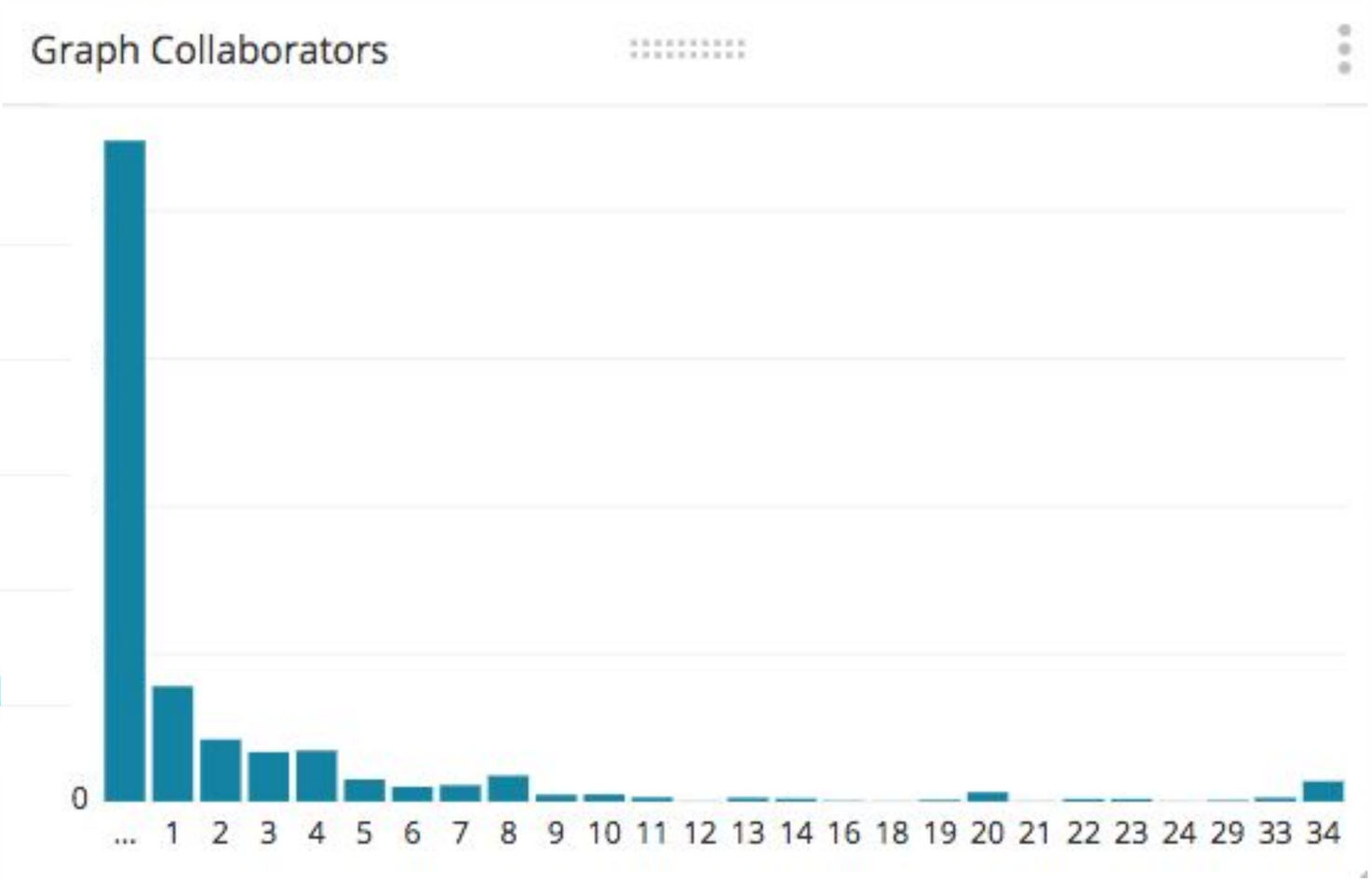


Boxplots in the Wild

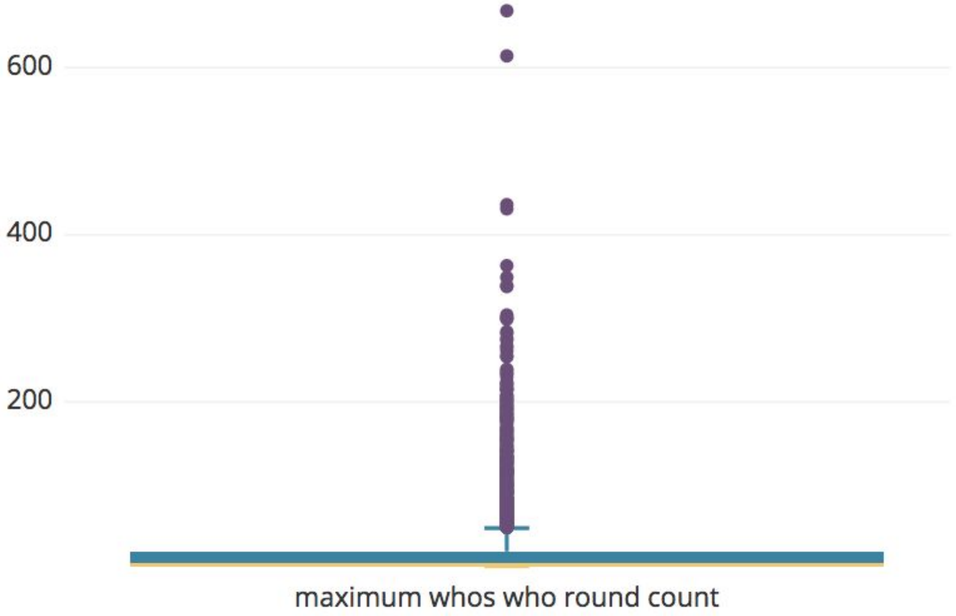
Anywhere an average can go...

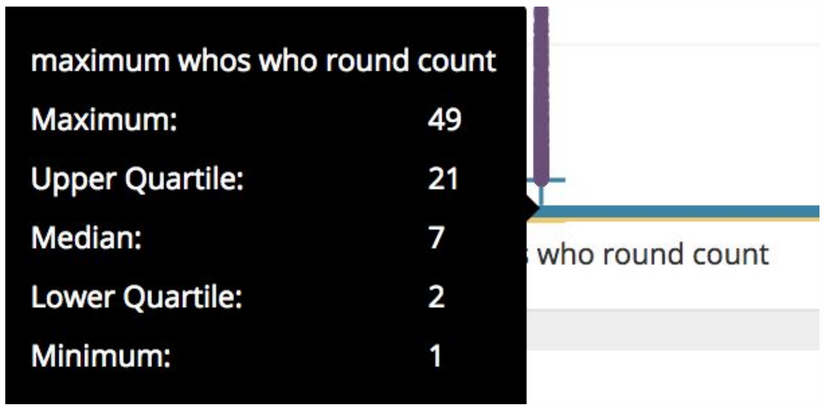
Examples from Pingboard



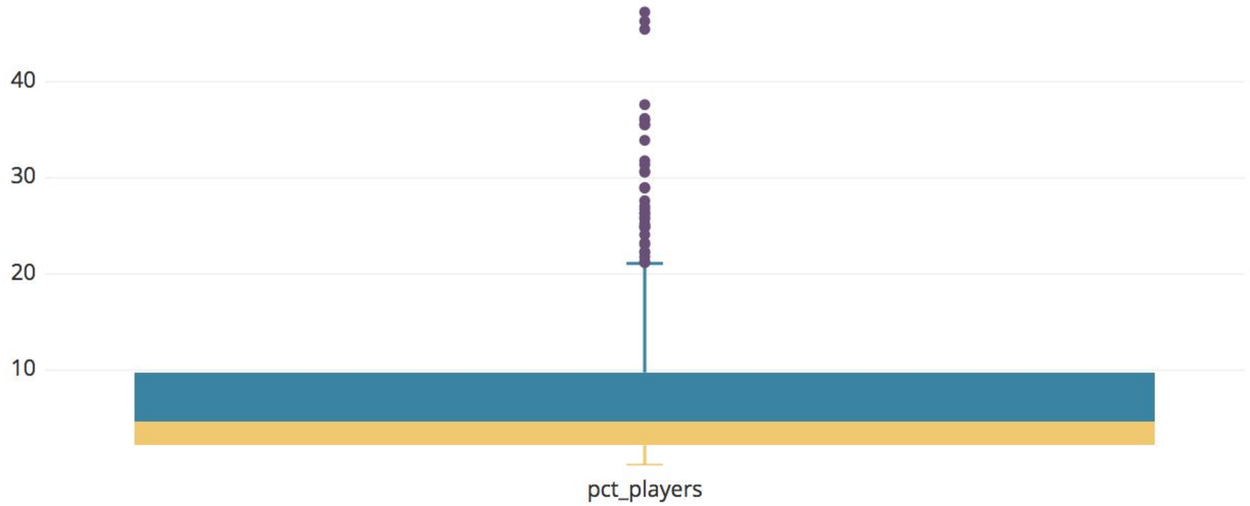


How Often do People Play?



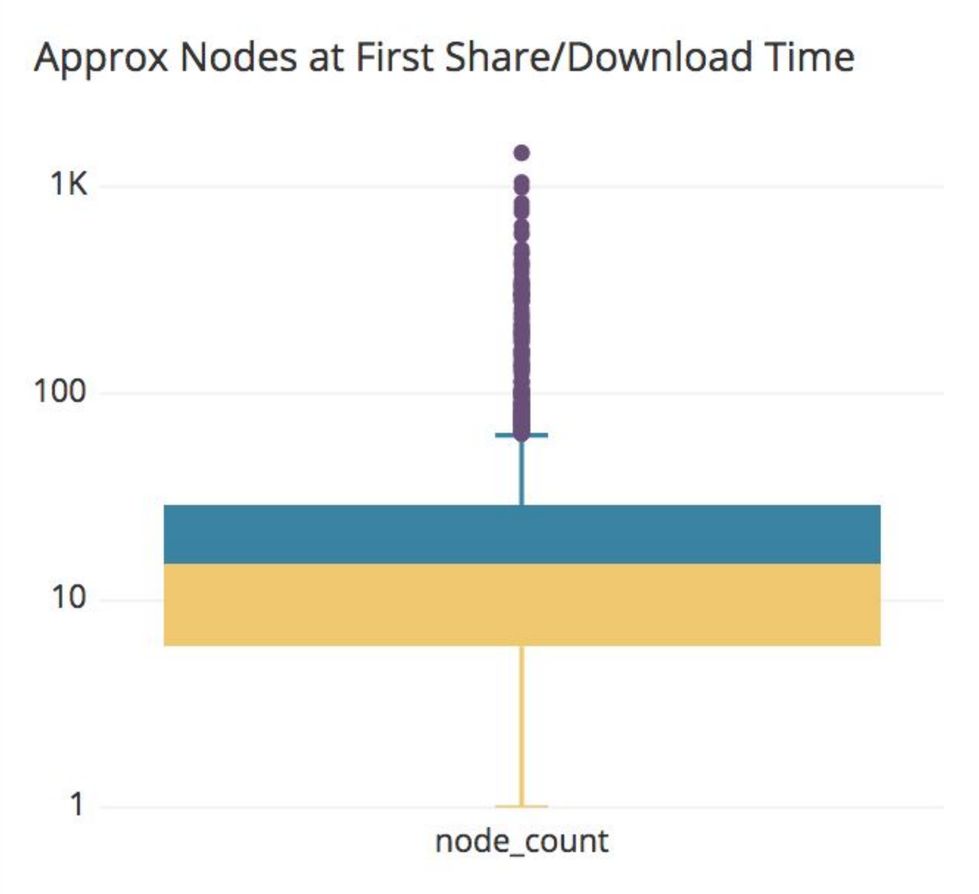


What % of each company plays if at least 2 play?

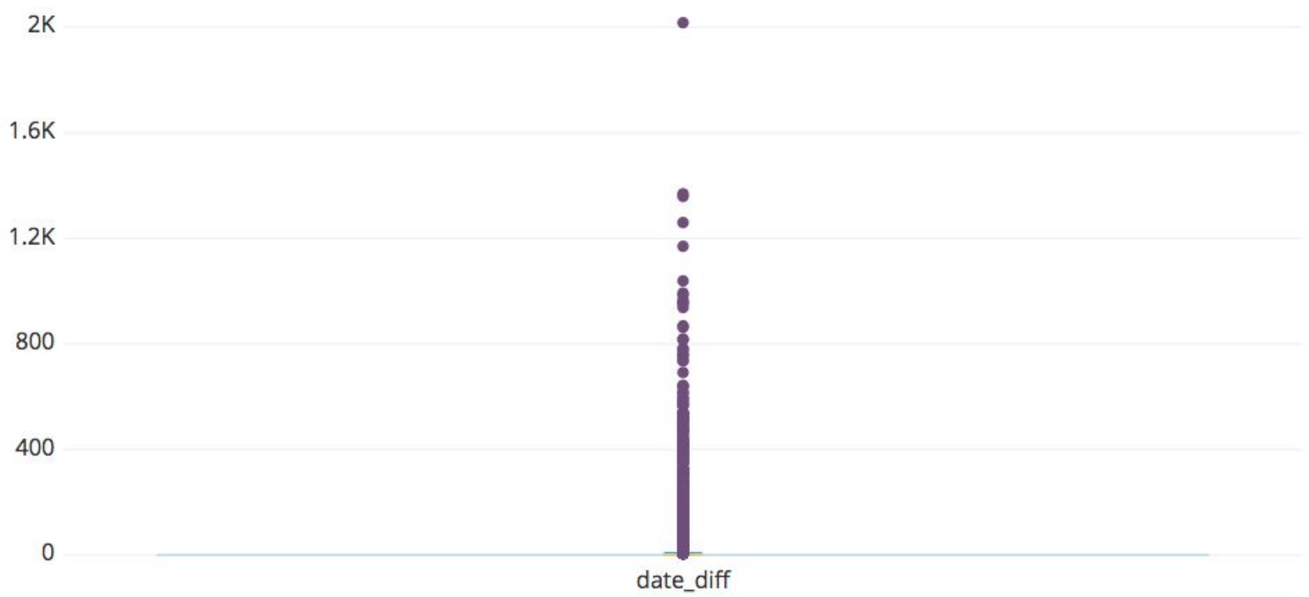


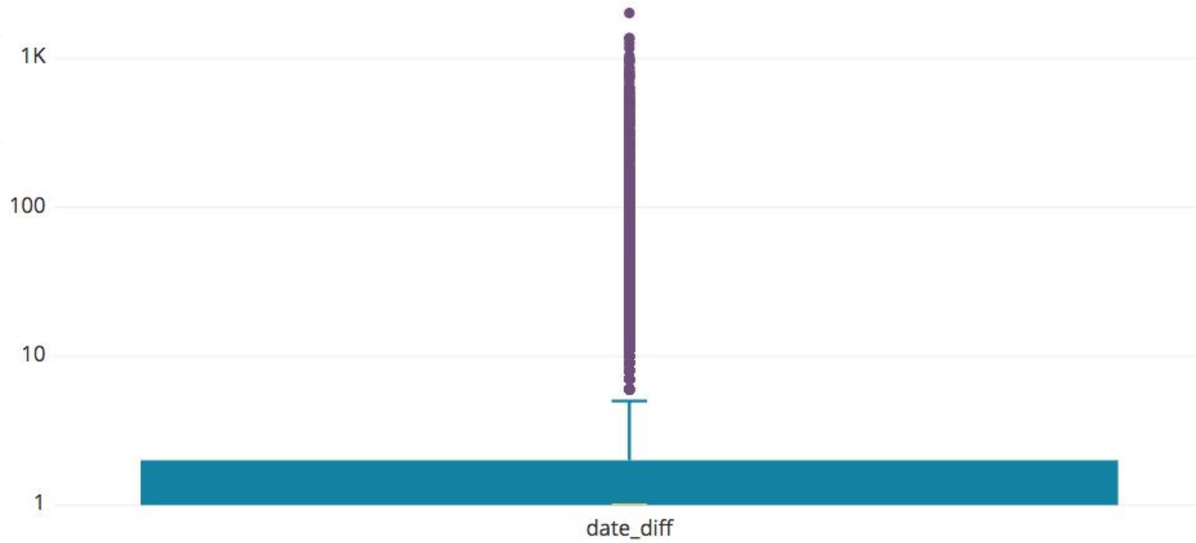
Examples from Pingboard

Logarithmic Scale!



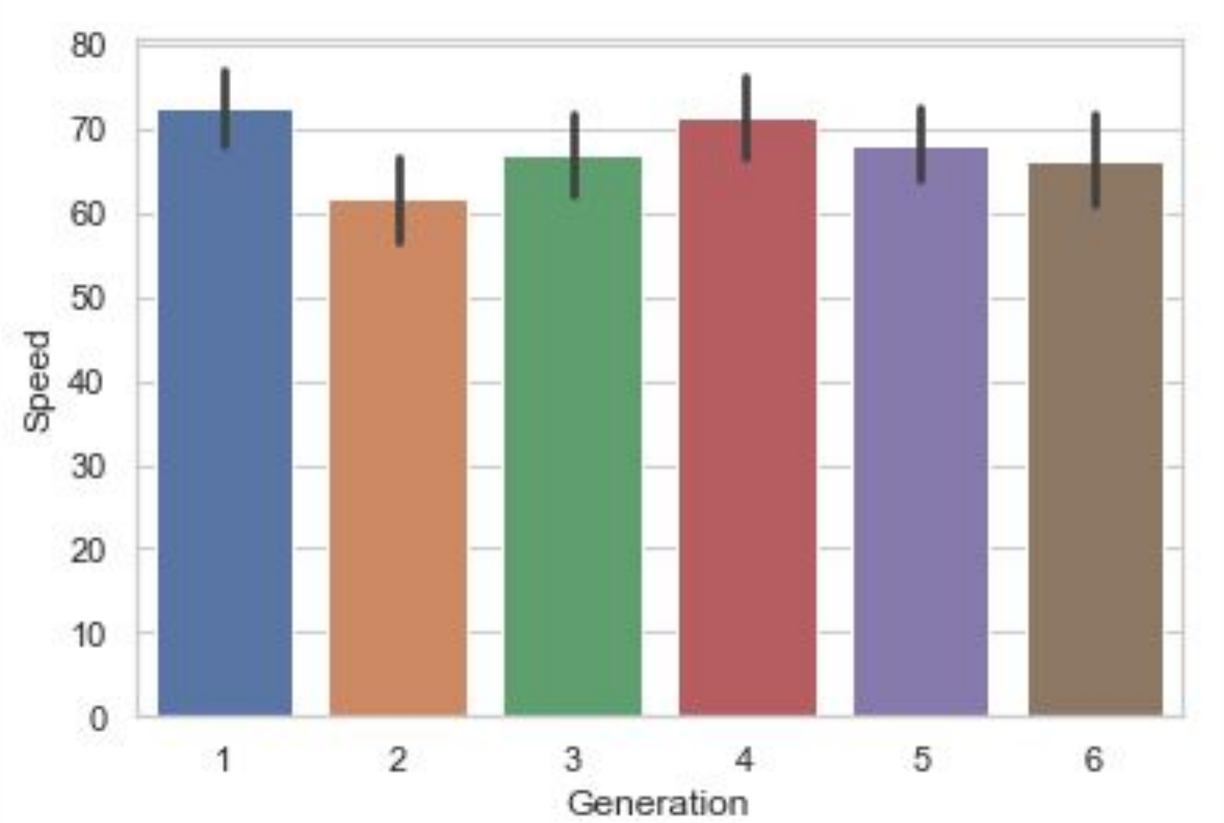
Examples from Pingboard

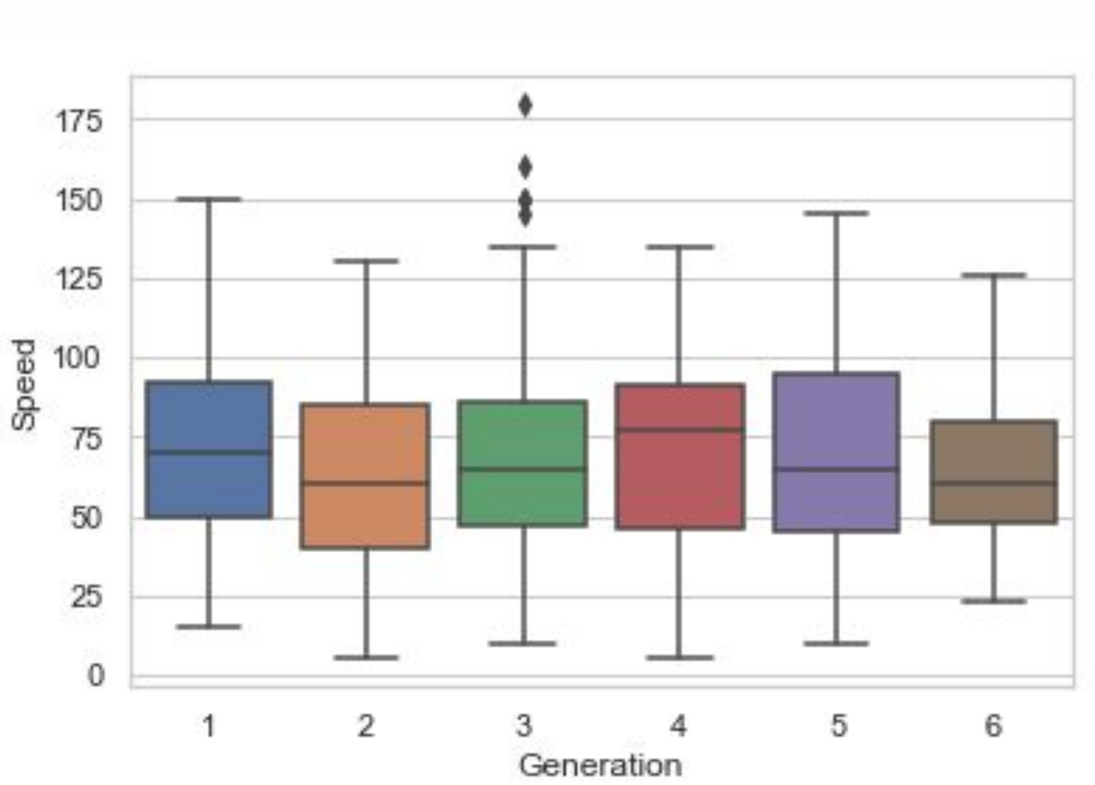




Not great for highly skewed data,
even with the logarithmic scale

Bar plots vs Boxplots





Boxplots aren’t Perfect

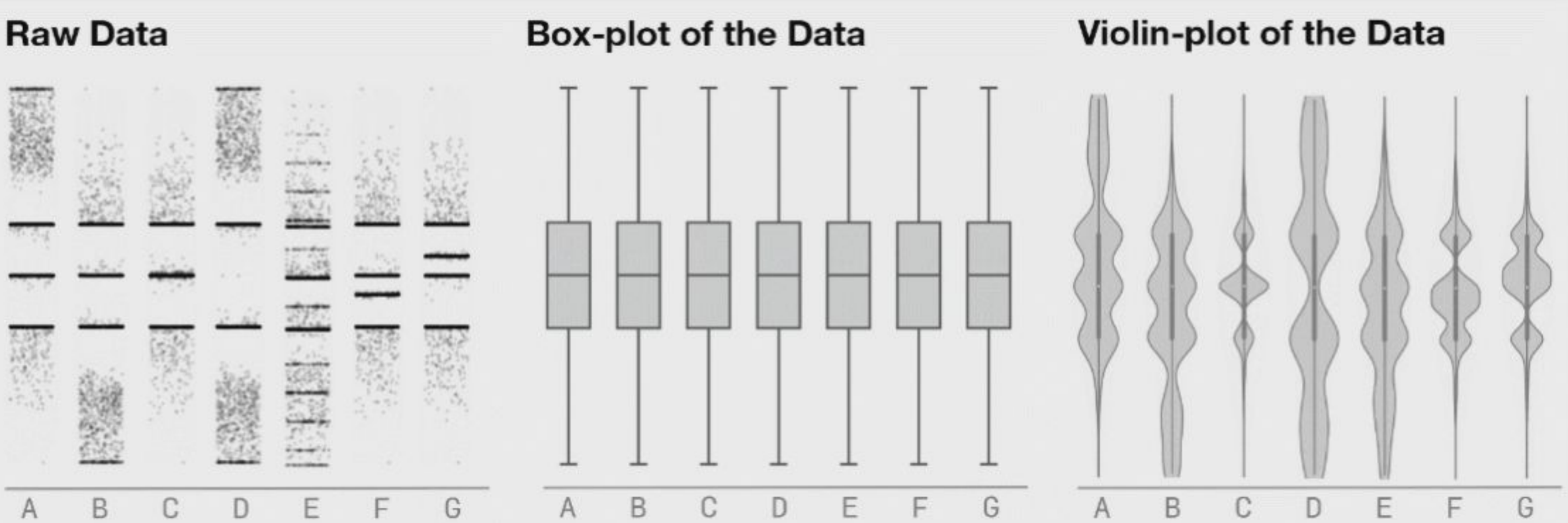
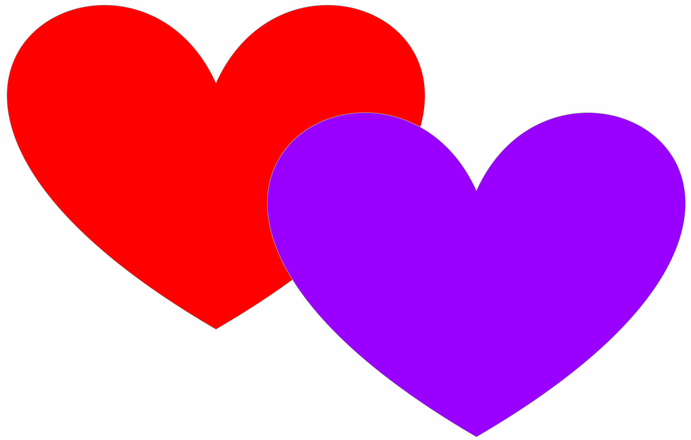


Fig 8. Seven distributions of data, shown as raw data points (or strip-plots), as box-plots, and as violin-plots.

https://www.autodeskresearch.com/publications/samestats
via
https://timo-roettger.weebly.com/thinking-outside-the-boxplot/think-outside-the-boxplot

Boxplots!

- give you a measure of central tendancy but also so much more
- slice and dice your data
- look for outliers
- can be made in a variety of tools



A Love Letter to the Boxplot

Because one number is never enough to describe a data set