Histogram

Histogram on a continuous variable can be accomplished using either geom\_bar() or geom\_histogram(). When using geom\_histogram(), you can control the number of bars using the bins option. Else, you can set the range covered by each bin using binwidth. The value of binwidth is on the same scale as the continuous variable on which histogram is built. Since, geom\_histogram gives facility to control both number of bins as well as binwidth, it is the preferred option to create histogram on continuous variables

**library**(ggplot2)

**theme\_set**(**theme\_classic**())

*# Histogram on a Continuous (Numeric) Variable*

g <- **ggplot**(mpg, **aes**(displ)) + **scale\_fill\_brewer**(palette = "Spectral")

g + **geom\_histogram**(**aes**(fill=class),

binwidth = .1,

col="black",

size=.1) + *# change binwidth*

**labs**(title="Histogram with Auto Binning",

subtitle="Engine Displacement across Vehicle Classes")

g + **geom\_histogram**(**aes**(fill=class),

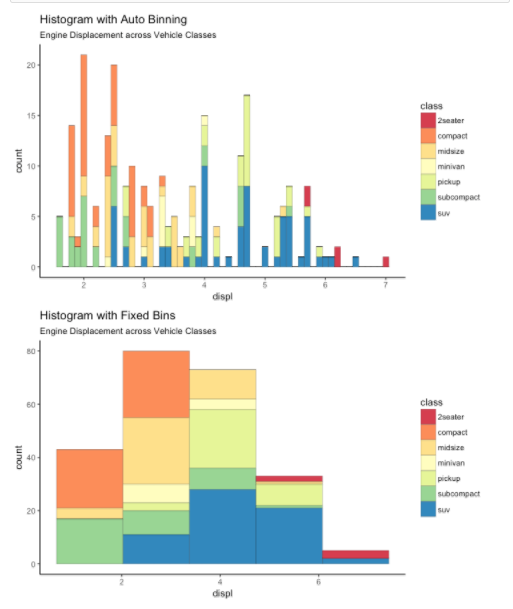
bins=5,

col="black",

size=.1) + *# change number of bins*

**labs**(title="Histogram with Fixed Bins",

subtitle="Engine Displacement across Vehicle Classes")



Histogram on a categorical variable

Histogram on a categorical variable would result in a frequency chart showing bars for each category. By adjusting width, you can adjust the thickness of the bars.

**library**(ggplot2)

**theme\_set**(**theme\_classic**())

*# Histogram on a Categorical variable*

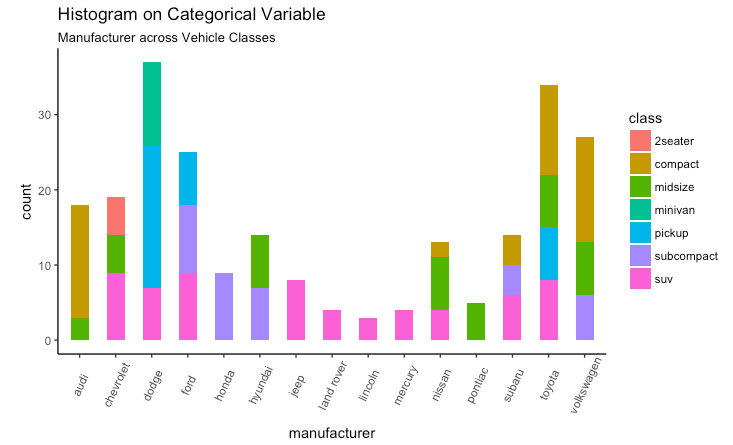
g <- **ggplot**(mpg, **aes**(manufacturer))

g + **geom\_bar**(**aes**(fill=class), width = 0.5) +

**theme**(axis.text.x = **element\_text**(angle=65, vjust=0.6)) +

**labs**(title="Histogram on Categorical Variable",

subtitle="Manufacturer across Vehicle Classes")



### **Density plot**

**library**(ggplot2)

**theme\_set**(**theme\_classic**())

*# Plot*

g <- **ggplot**(mpg, **aes**(cty))

g + **geom\_density**(**aes**(fill=**factor**(cyl)), alpha=0.8) +

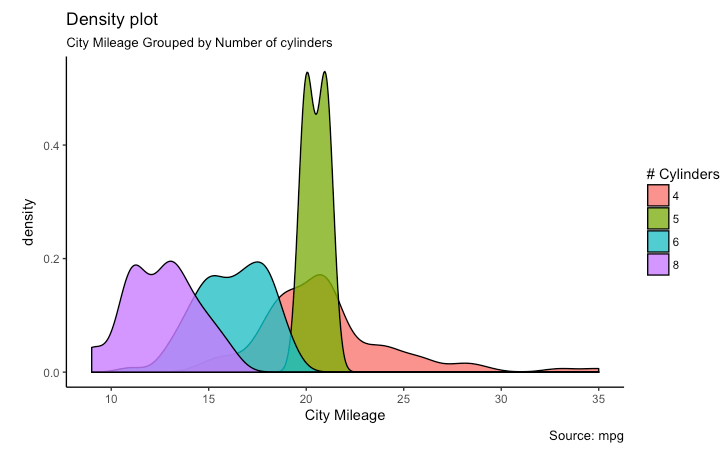
**labs**(title="Density plot",

subtitle="City Mileage Grouped by Number of cylinders",

caption="Source: mpg",

x="City Mileage",

fill="# Cylinders")



Box Plot

Box plot is an excellent tool to study the distribution. It can also show the distributions within multiple groups, along with the median, range and outliers if any.

The dark line inside the box represents the median. The top of box is 75%ile and bottom of box is 25%ile. The end points of the lines (aka whiskers) is at a distance of 1.5\*IQR, where IQR or Inter Quartile Range is the distance between 25th and 75th percentiles. The points outside the whiskers are marked as dots and are normally considered as extreme points.

Setting varwidth=T adjusts the width of the boxes to be proportional to the number of observation it contains.

**library**(ggplot2)

**theme\_set**(**theme\_classic**())

*# Plot*

g <- **ggplot**(mpg, **aes**(class, cty))

g + **geom\_boxplot**(varwidth=T, fill="plum") +

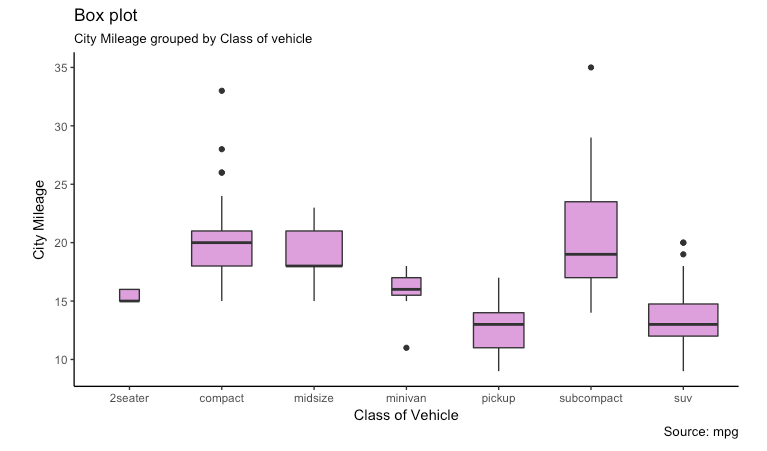
**labs**(title="Box plot",

subtitle="City Mileage grouped by Class of vehicle",

caption="Source: mpg",

x="Class of Vehicle",

y="City Mileage")



**library**(ggthemes)

g <- **ggplot**(mpg, **aes**(class, cty))

g + **geom\_boxplot**(**aes**(fill=**factor**(cyl))) +

**theme**(axis.text.x = **element\_text**(angle=65, vjust=0.6)) +

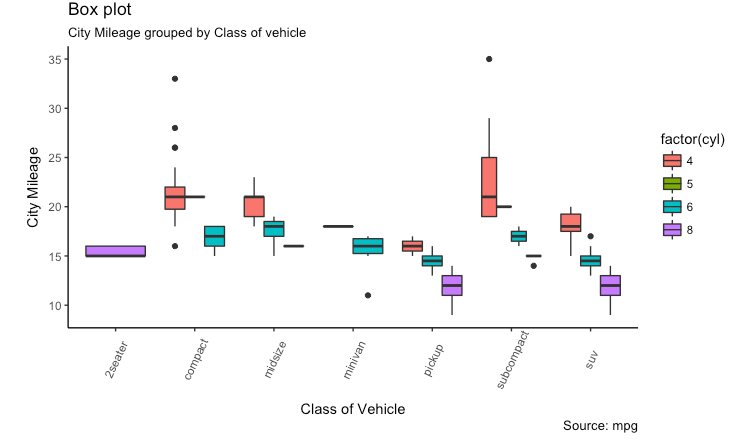
**labs**(title="Box plot",

subtitle="City Mileage grouped by Class of vehicle",

caption="Source: mpg",

x="Class of Vehicle",

y="City Mileage")



### **Dot + Box Plot**

On top of the information provided by a box plot, the dot plot can provide more clear information in the form of summary statistics by each group. The dots are staggered such that each dot represents one observation. So, in below chart, the number of dots for a given manufacturer will match the number of rows of that manufacturer in source data.

**library**(ggplot2)

**theme\_set**(**theme\_bw**())

*# plot*

g <- **ggplot**(mpg, **aes**(manufacturer, cty))

g + **geom\_boxplot**() +

**geom\_dotplot**(binaxis='y',

stackdir='center',

dotsize = .5,

fill="red") +

**theme**(axis.text.x = **element\_text**(angle=65, vjust=0.6)) +

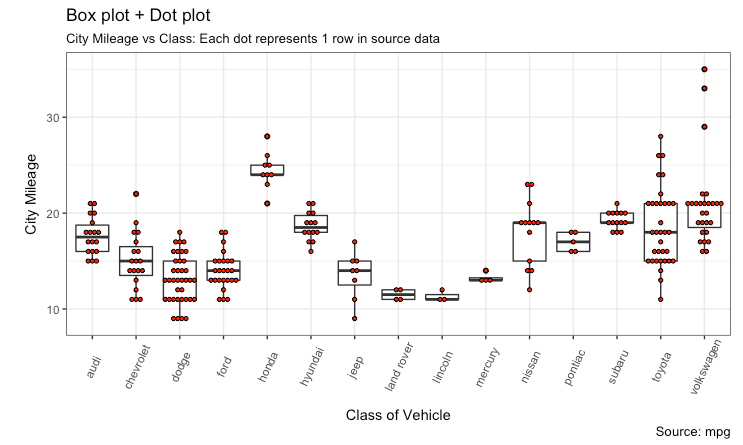
**labs**(title="Box plot + Dot plot",

subtitle="City Mileage vs Class: Each dot represents 1 row in source data",

caption="Source: mpg",

x="Class of Vehicle",

y="City Mileage")



Tufte Boxplot

Tufte box plot, provided by ggthemes package is inspired by the works of Edward Tufte. Tufte’s Box plot is just a box plot made minimal and visually appealing.

**library**(ggthemes)

**library**(ggplot2)

**theme\_set**(**theme\_tufte**()) *# from ggthemes*

*# plot*

g <- **ggplot**(mpg, **aes**(manufacturer, cty))

g + **geom\_tufteboxplot**() +

**theme**(axis.text.x = **element\_text**(angle=65, vjust=0.6)) +

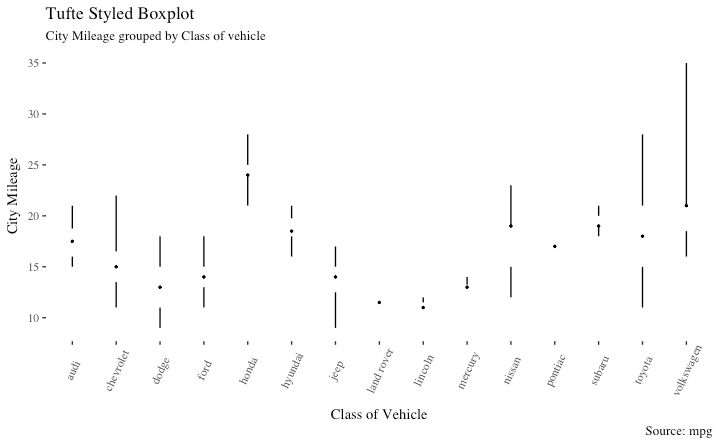
**labs**(title="Tufte Styled Boxplot",

subtitle="City Mileage grouped by Class of vehicle",

caption="Source: mpg",

x="Class of Vehicle",

y="City Mileage")



Violin Plot

A violin plot is similar to box plot but shows the density within groups. Not much info provided as in boxplots. It can be drawn using geom\_violin().

**library**(ggplot2)

**theme\_set**(**theme\_bw**())

*# plot*

g <- **ggplot**(mpg, **aes**(class, cty))

g + **geom\_violin**() +

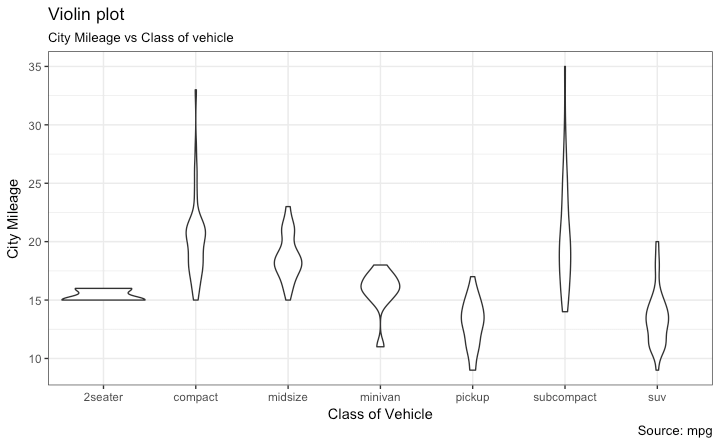
**labs**(title="Violin plot",

subtitle="City Mileage vs Class of vehicle",

caption="Source: mpg",

x="Class of Vehicle",

y="City Mileage")



Population Pyramid

Population pyramids offer a unique way of visualizing how much population or what percentage of population fall under a certain category. The below pyramid is an excellent example of how many users are retained at each stage of a email marketing campaign funnel.

**library**(ggplot2)

**library**(ggthemes)

**options**(scipen = 999) *# turns of scientific notations like 1e+40*

*# Read data*

email\_campaign\_funnel <- **read.csv**("https://raw.githubusercontent.com/selva86/datasets/master/email\_campaign\_funnel.csv")

*# X Axis Breaks and Labels*

brks <- **seq**(-15000000, 15000000, 5000000)

lbls = **paste0**(**as.character**(**c**(**seq**(15, 0, -5), **seq**(5, 15, 5))), "m")

*# Plot*

**ggplot**(email\_campaign\_funnel, **aes**(x = Stage, y = Users, fill = Gender)) + *# Fill column*

**geom\_bar**(stat = "identity", width = .6) + *# draw the bars*

**scale\_y\_continuous**(breaks = brks, *# Breaks*

labels = lbls) + *# Labels*

**coord\_flip**() + *# Flip axes*

**labs**(title="Email Campaign Funnel") +

**theme\_tufte**() + *# Tufte theme from ggfortify*

**theme**(plot.title = **element\_text**(hjust = .5),

axis.ticks = **element\_blank**()) + *# Centre plot title*

**scale\_fill\_brewer**(palette = "Dark2") *# Color palette*

