

# Basic R Visualization

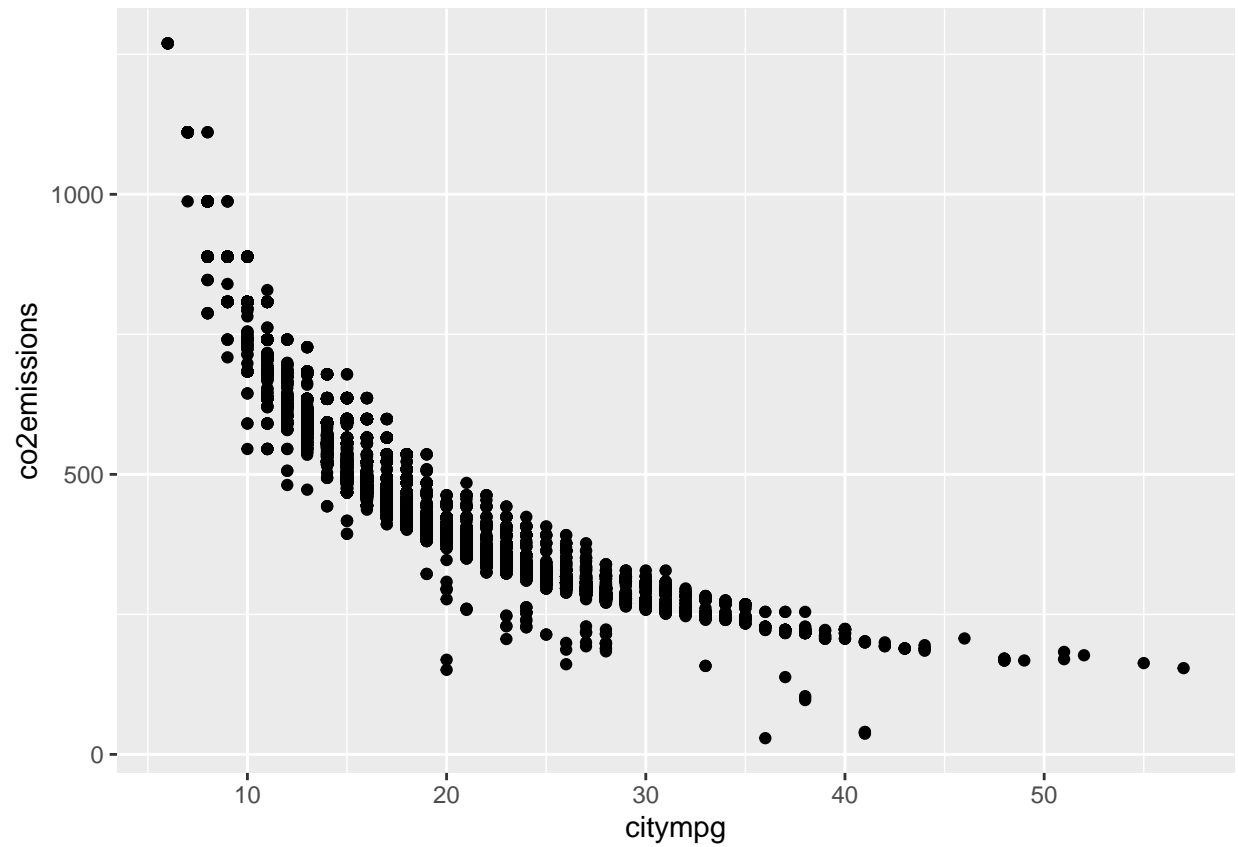
2023-10-01

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.0      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.1      v tibble    3.1.8
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the ]8;;http://conflicted.r-lib.org/conflicted-package]8;; to force all conflicts to become errors
```

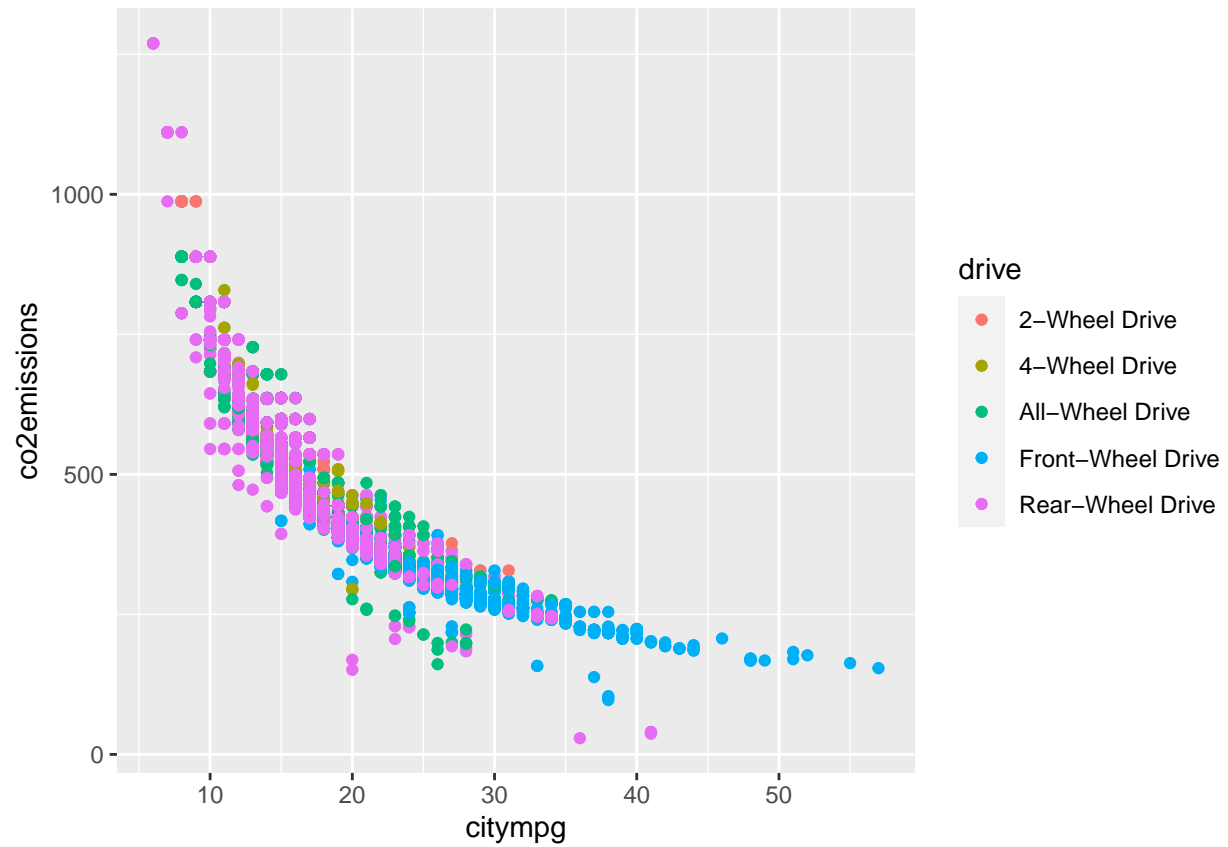
```
library(dplyr)
vehicles <- read_csv("https://s3.amazonaws.com/itao-30230/vehicles.csv",
                    col_types="inincicccici")
vehicles <- vehicles %>%
  mutate(class=as.factor(class), drive=as.factor(drive), make=as.factor(make),
         transmissiontype=as.factor(transmissiontype), citympg = as.numeric(citympg))
view(vehicles)
```

```
#1 Part 1 scatter plot
ggplot(data = vehicles) +
  geom_point(mapping = aes(x = citympg, y =co2emissions))
```



*#2 Part B. Scatter plot with colors*

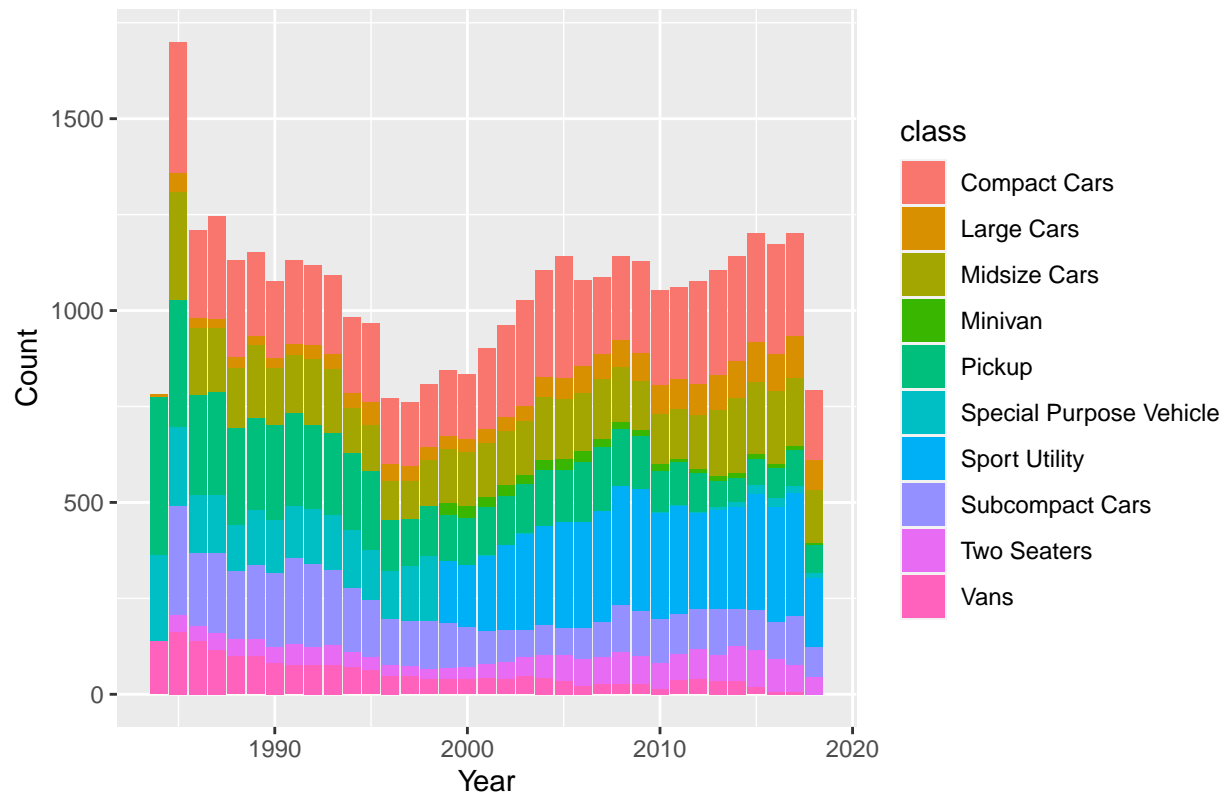
```
ggplot(data = vehicles) +  
  geom_point(mapping = aes(x=citympg, y = co2emissions, color = drive))
```



*#3 Part C. Stacked bar chart with the number of vehicles tested each year*

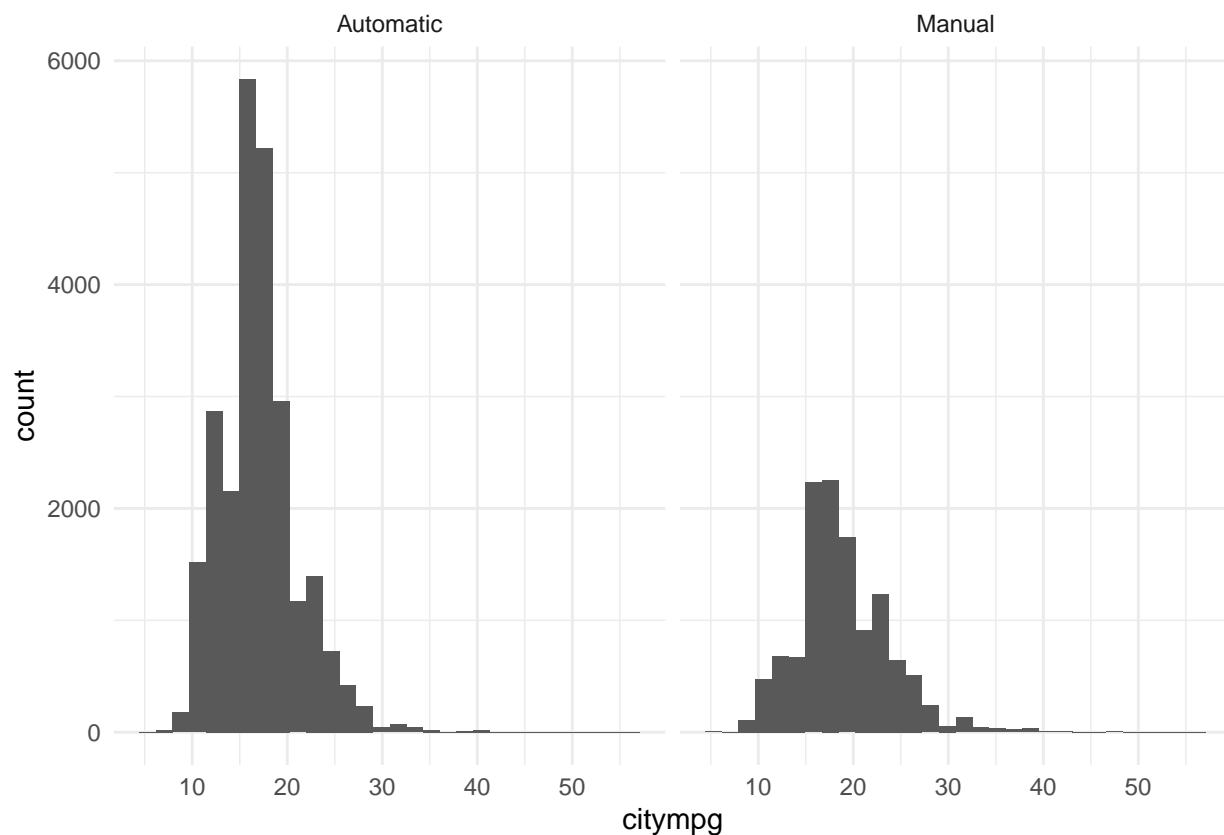
```
ggplot(vehicles, aes(fill = class, x = year)) +
  geom_bar(position = "stack") +
  labs(
    title = "Number of Vehicles Tested by Year and Vehicle Class",
    x = "Year",
    y = "Count"
  )
```

Number of Vehicles Tested by Year and Vehicle Class



```
#4 Part D. Histogram
ggplot(data = vehicles) +
  geom_histogram(mapping = aes(x = citympg)) +
  facet_wrap(~transmissiontype) +
  theme_minimal()
```

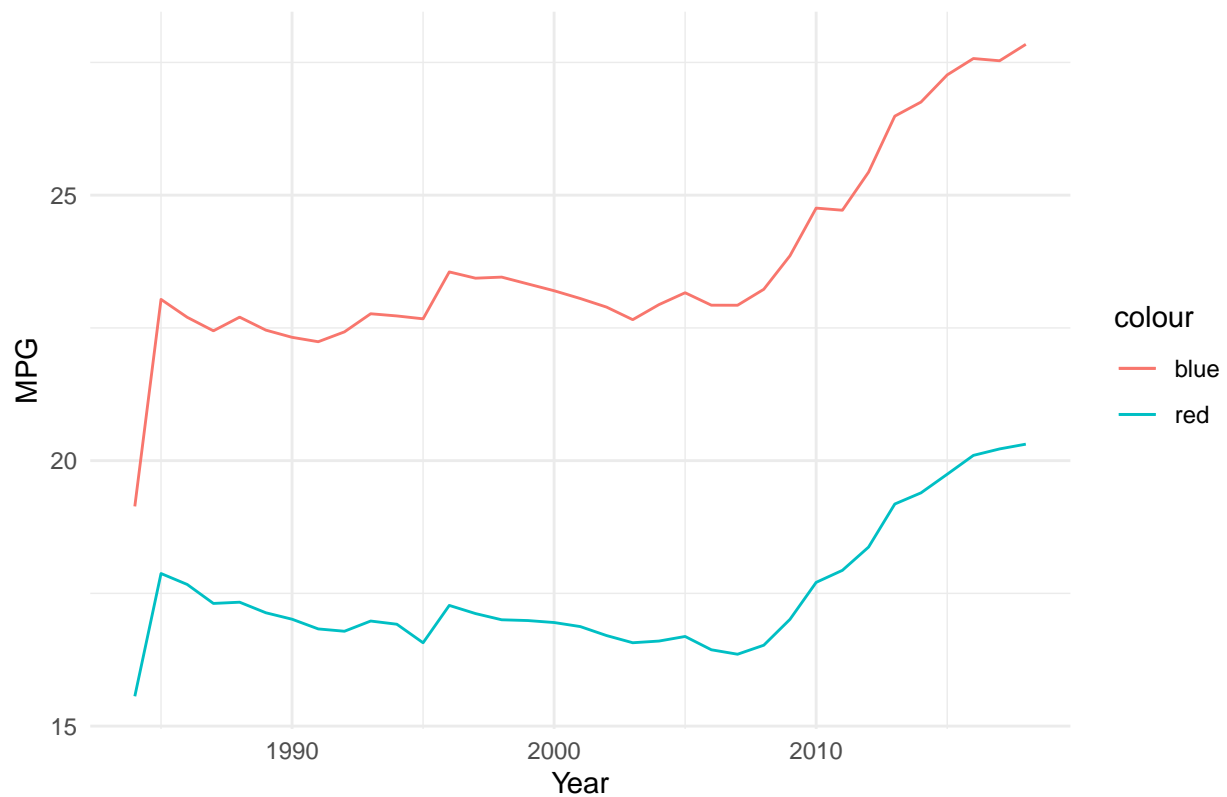
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
#Part 2.
#Part A. Summary statistics
tab <- vehicles %>%
  group_by(class) %>%
  summarize(minimum_mpg = min(citympg), maximum_mpg = max(citympg), mean_mpg = round(mean(citympg)), med
view(tab)
```

```
#Part B.
vehicles %>%
  group_by(year) %>%
  summarize(mean_citympg = mean(citympg, na.rm = TRUE),
            mean_highwaympg = mean(highwaympg, na.rm = TRUE)) %>%
  ggplot() +
  geom_line(aes(x = year, y = mean_citympg, color = 'red')) +
  geom_line(aes(x = year, y = mean_highwaympg, color = 'blue')) +
  labs(
    title = "Change in Average City vs. Highway MPG Over Time",
    x = "Year",
    y = "MPG"
  ) +
  theme_minimal()
```

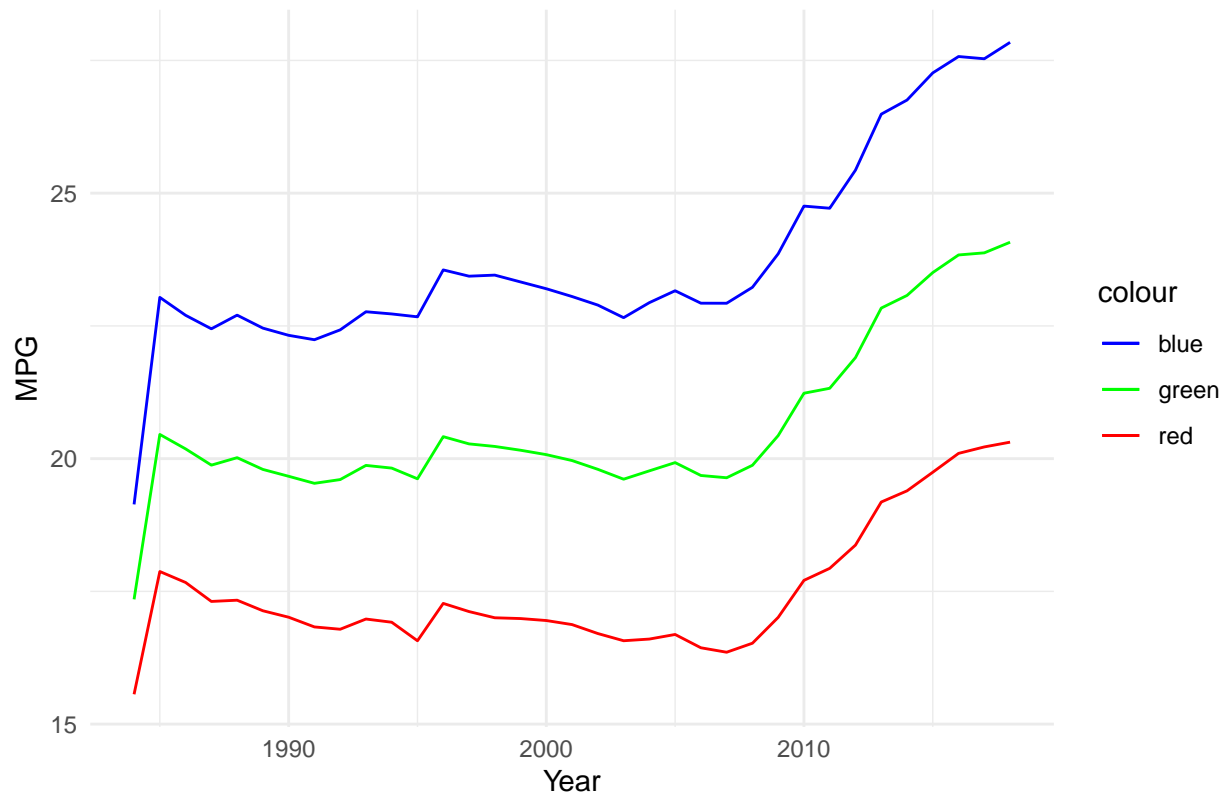
Change in Average City vs. Highway MPG Over Time



*#Part C. Line with average*

```
vehicles %>%
  group_by(year) %>%
  summarize(
    mean_citympg = mean(citympg, na.rm = TRUE),
    mean_highwaympg = mean(highwaympg, na.rm = TRUE),
    average_overallmpg = mean((highwaympg + citympg) / 2, na.rm = TRUE)
  ) %>%
  ggplot() +
  geom_line(aes(x = year, y = mean_citympg, color = 'red')) +
  geom_line(aes(x = year, y = mean_highwaympg, color = 'blue')) +
  geom_line(aes(x = year, y = average_overallmpg, color = 'green')) +
  labs(
    title = "Change in Average City vs. Highway vs Overall MPG Over Time",
    x = "Year",
    y = "MPG"
  ) +
  scale_color_manual(values = c('red' = 'red', 'blue' = 'blue', 'green' = 'green')) +
  theme_minimal()
```

## Change in Average City vs. Highway vs Overall MPG Over Time



*#Part D*

```
vehicles %>%
  group_by(year, drive) %>%
  summarize(
    mean_citympg = mean(citympg, na.rm = TRUE),
    mean_highwaympg = mean(highwaympg, na.rm = TRUE),
    average_overallmpg = mean((highwaympg + citympg) / 2, na.rm = TRUE)
  ) %>%
  ggplot() +
  geom_line(aes(x = year, y = mean_citympg, color = 'red')) +
  geom_line(aes(x = year, y = mean_highwaympg, color = 'blue')) +
  geom_line(aes(x = year, y = average_overallmpg, color = 'green')) +
  facet_wrap(~drive) +
  labs(
    title = "Change in Average City vs. Highway vs Overall MPG Over Time",
    x = "Year",
    y = "MPG"
  ) +
  theme_minimal()
```

## 'summarise()' has grouped output by 'year'. You can override using the  
## '.groups' argument.