

LV5

Single Variable Visualisation

Michael Scheidl

06.10.2021

Contents

Libraries	2
1 Overview	2
2 Plots for a single Variable - bar plot	2
2.1 Classic Bar Chart	2
2.2 Stacked Bar Chart	3
2.3 Pie Chart	3
2.4 Graphs with two categorial variables	4
2.5 Stacked pie Charts - two categorial variables	5
2.6 Grouped bar charts	7
3 Exercise	8
4 Once continuous variable	9
4.1 Histogram	9
4.2 Boxplot	9
4.3 Erklärende Variable sollten eig PS sein	11
4.4 Scatter plot using labels	12
5 Using ggrepel package	12
6 One categorial and one numeric variable	13
6.1 Boxplot	13
6.2 Violin plot	14
6.3 Facetted Plot	14
6.4 Exercise 2	15

Libraries

```
library(tidyverse)
```

1 Overview

- Single Variable Visualisation
- Pie Charts
- Bar Charts
- Stacked Bar Charts
- Density Plot
- Boxplot
- Violin Plot
- Mosaic Plot

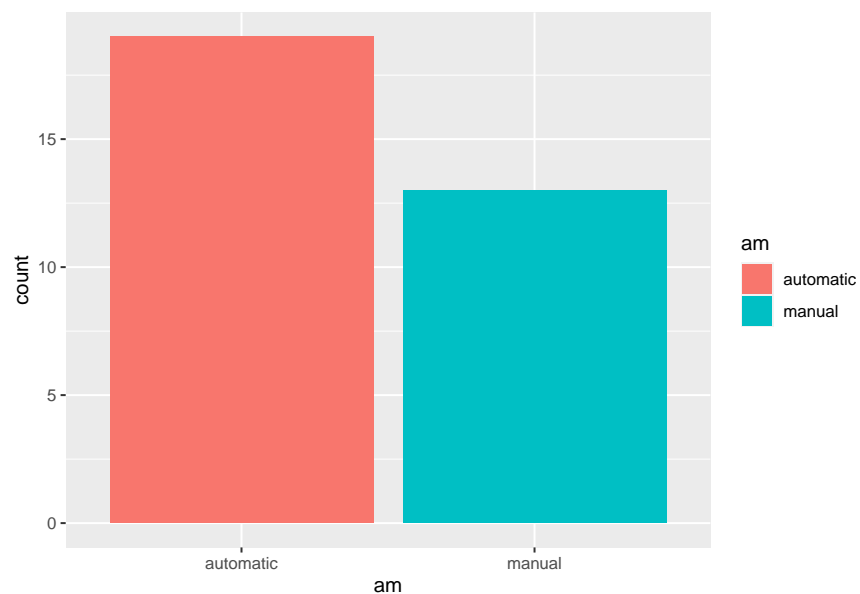
2 Plots for a single Variable - bar plot

Create Dataset

```
mtcars %>%  
  mutate(  
    am = ifelse(am == 1, "manual", "automatic"),  
    vs = ifelse(vs == 1, "straight", "v-shaped")  
  ) -> data
```

2.1 Classic Bar Chart

```
data %>% ggplot(aes(x = am, fill = am)) + geom_bar()
```



Interpretation: - shows the number (count) of cars with automatic and manual shifting - data set contains more cars with automatic shifting than with manual shifting of gears

2.2 Stacked Bar Chart

```
data %>% ggplot(aes(x = "", fill = am)) +  
  geom_bar(position = "fill", width = 0.5) +  
  xlab("") +  
  ylab("Proportion") +  
  scale_x_discrete(breaks = NULL)
```

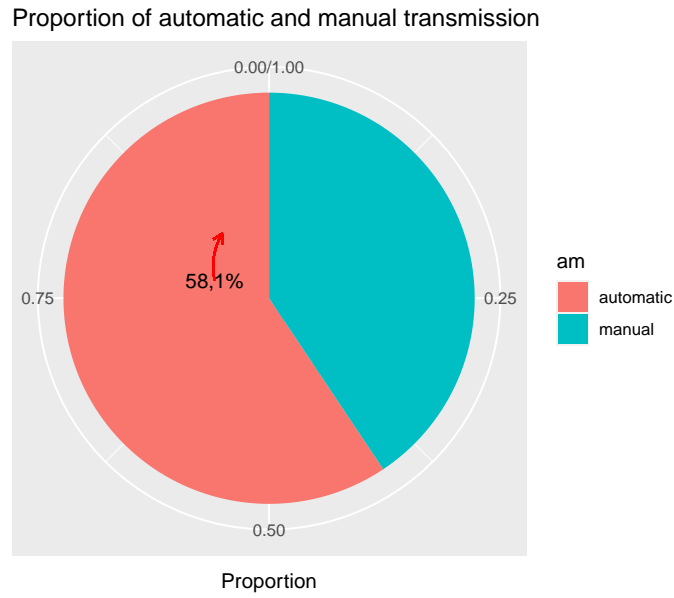


Interpretation:

- the graph shows the proportion of automatic to manual gear shifting in the car data set

2.3 Pie Chart

```
ggplot(data, aes(x = 1, fill=am)) +  
  geom_bar(position = "fill") +  
  coord_polar("y") +  
  xlab("") +  
  ylab("Proportion") +  
  ggtitle("Proportion of automatic and manual transmission") +  
  scale_x_continuous(breaks = NULL) +  
  #annotate("text", 0.3, 0.3, label="58,1%") +  
  geom_segment(aes(x = 0.8, y = 0.8, xend = 0.9, yend = 0.9),  
    arrow = arrow(length = unit(0.2, "cm")),  
    col = "red") +  
  annotate("text", 0.8, 0.8, label="58,1%")
```

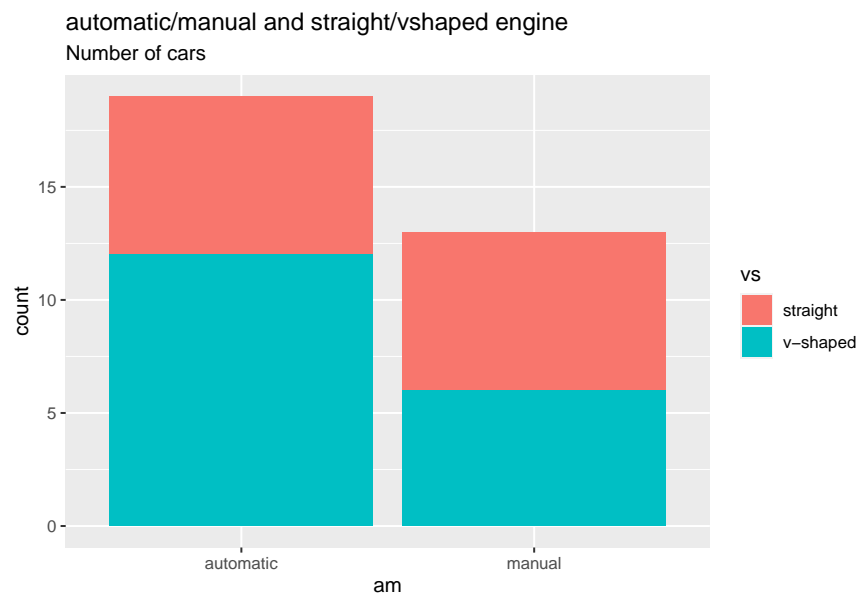


Interpretation:

- Different form of previous plot
- shows that 58.1% of the cars in the dataset are automatic transmission

2.4 Graphs with two categorical variables

```
data %>% ggplot(aes(x = am, fill=vs)) +
  geom_bar() +
  ggtitle("automatic/manual and straight/vshaped engine", subtitle="Number of cars")
```

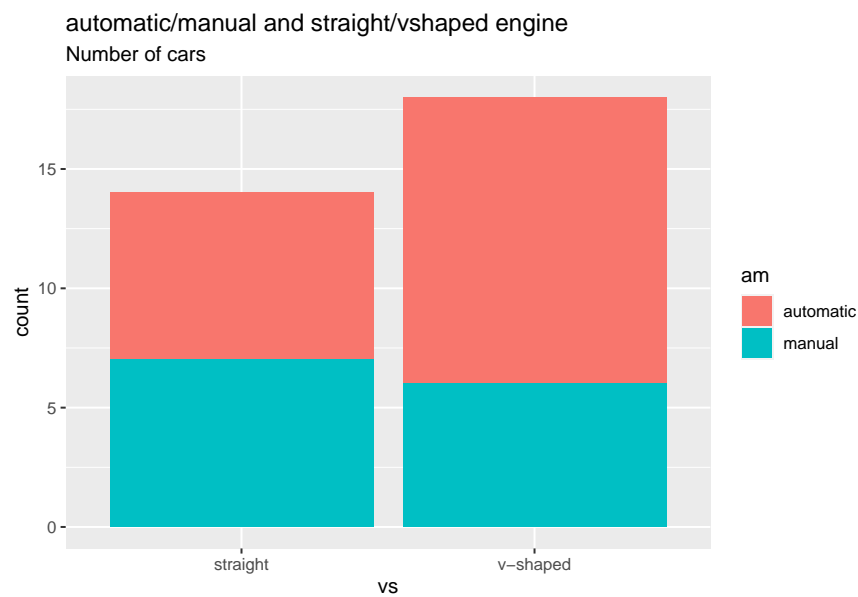


Interpretation:

- Number of cars with automatic gear shifting is higher than cars with manual gear shifting
- For automatic cars the v-shaped engines dominate (around 2/3)
- Manual transmission has approximately equal straight and vshaped engines

2.4.1 Alternativ Switch variables

```
data %>% ggplot(aes(x = vs, fill=am)) +  
  geom_bar() +  
  ggtitle("automatic/manual and straight/vshaped engine", subtitle="Number of cars")
```

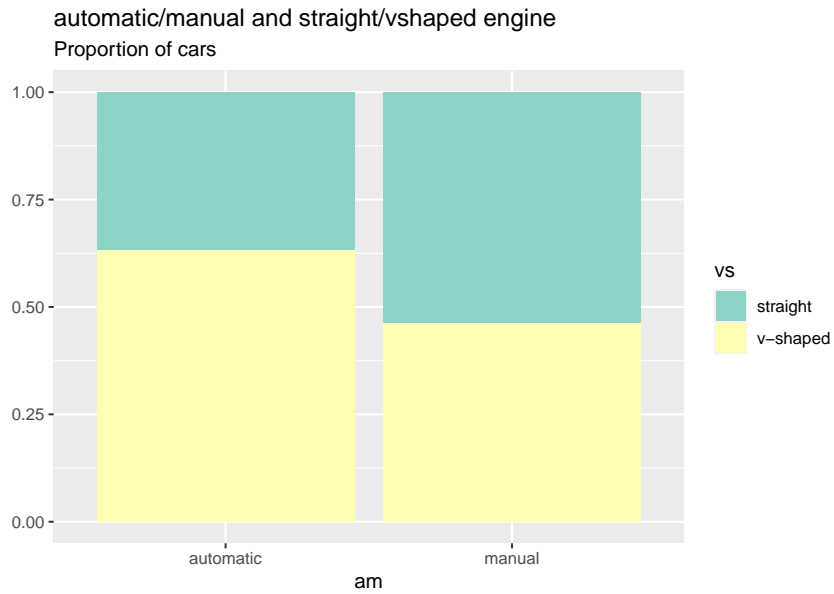


Interpretation:

- Doesn't change much for interpretation

2.5 Stacked pie Charts - two categorical variables

```
data %>% ggplot(aes(x = am, fill=vs)) +  
  geom_bar(position = "fill") +  
  ggtitle("automatic/manual and straight/vshaped engine", subtitle="Proportion of cars") +  
  ylab("") +  
  scale_fill_brewer(type = "qual", palette = 8)
```

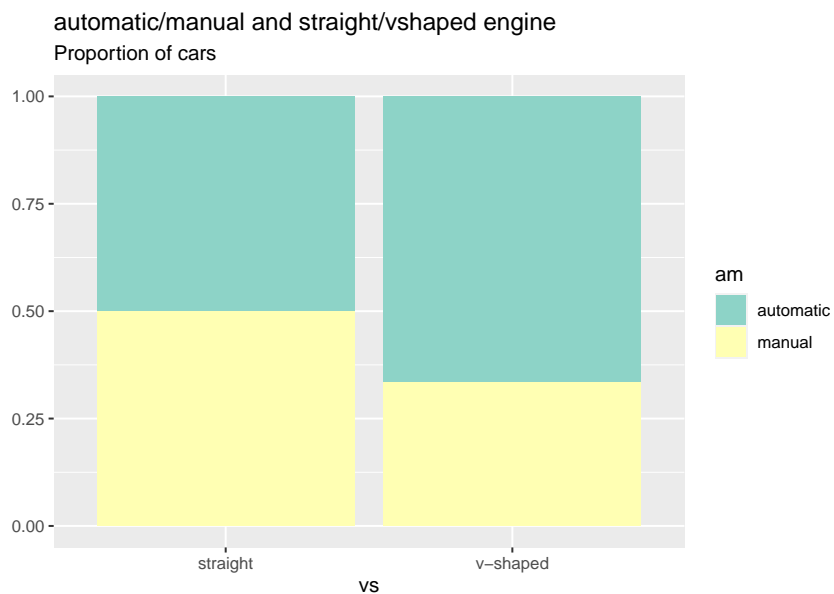


Interpretation:

- Share of straight engines is lower with automatic gearbox than with manual gear boxes
- That more cars are automatic is no longer visible in the graph

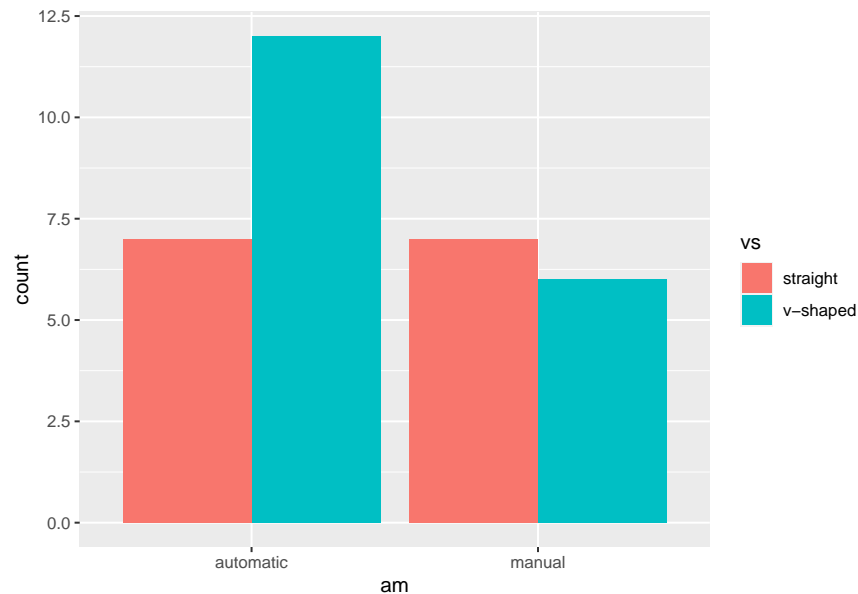
2.5.1 Alternative - switch variables

```
data %>% ggplot(aes(x = vs, fill=am)) +
  geom_bar(position = "fill") +
  ggtitle("automatic/manual and straight/vshaped engine", subtitle="Proportion of cars") +
  ylab("") +
  scale_fill_brewer(type = "qual", palette = 8)
```



2.6 Grouped bar charts

```
data %>% ggplot(aes(x = am , fill= vs)) +  
  geom_bar(position = "dodge")
```



Interpretation:

- Two groups are compared (automatic vs manual)
- In the group automatic the number of straight engines is lower than the number of v-shaped engines. In the group if manual shifting gear cars the opposite is the case.
- The number of straight engines appears to be equal in the group of automatic and manual cars

Check number of cars via data manipulation

```
data %>% group_by(am, vs) %>% summarise(n=n())
```

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

```
## # A tibble: 4 x 3  
## # Groups:   am [2]  
##   am      vs      n  
##   <chr>   <chr> <int>  
## 1 automatic straight    7  
## 2 automatic v-shaped   12  
## 3 manual   straight    7  
## 4 manual   v-shaped    6
```

Interpretation:

- Indeed the number of cars with straight engines is equal in the group of automatic and manual cars.

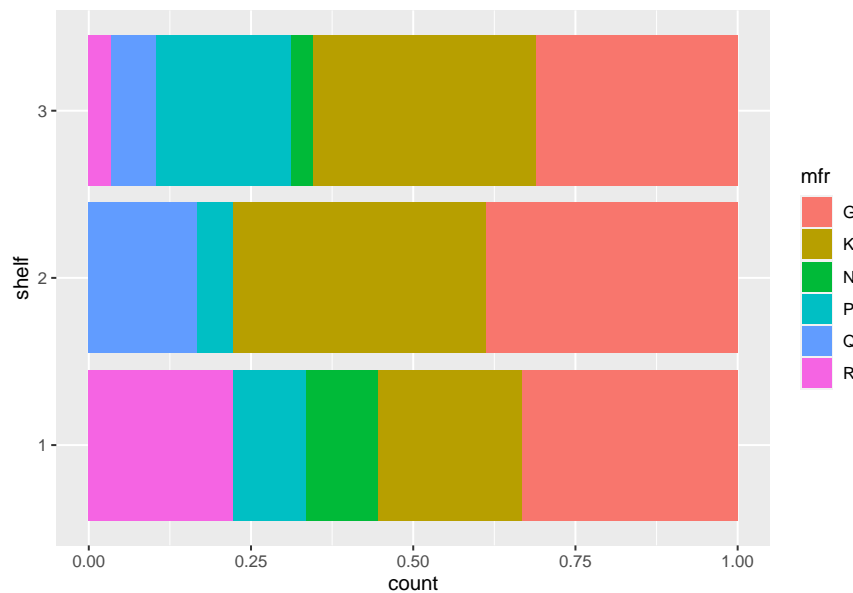
3 Exercise

Visualize the proportion of manufacturers among the shelves

```
data(UScereal, package = "MASS")
head(UScereal)
```

```
##               mfr calories  protein    fat  sodium  fibre
## 100% Bran      N 212.1212 12.121212 3.030303 393.9394 30.303030
## All-Bran      K 212.1212 12.121212 3.030303 787.8788 27.272727
## All-Bran with Extra Fiber K 100.0000 8.000000 0.000000 280.0000 28.000000
## Apple Cinnamon Cheerios G 146.6667 2.666667 2.666667 240.0000 2.000000
## Apple Jacks   K 110.0000 2.000000 0.000000 125.0000 1.000000
## Basic 4       G 173.3333 4.000000 2.666667 280.0000 2.666667
##               carbo  sugars shelf potassium vitamins
## 100% Bran      15.15152 18.18182    3 848.48485 enriched
## All-Bran      21.21212 15.15151    3 969.69697 enriched
## All-Bran with Extra Fiber 16.00000 0.00000    3 660.00000 enriched
## Apple Cinnamon Cheerios 14.00000 13.33333    1 93.33333 enriched
## Apple Jacks   11.00000 14.00000    2 30.00000 enriched
## Basic 4       24.00000 10.66667    3 133.33333 enriched
```

```
UScereal %>% mutate(shelf = as.factor(shelf)) %>% ggplot(aes(x = shelf, fill=mfr)) + geom_bar(position = "stack")
```

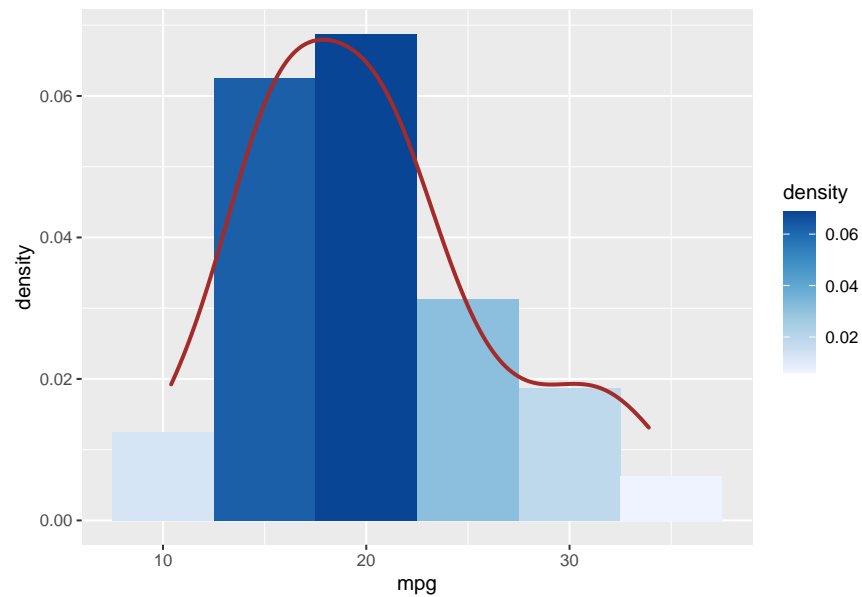


```
#UScereal %>% mutate(shelf = as.factor(shelf)) %>% ggplot(aes(x = mfr, fill=shelf)) + geom_bar(position = "stack")
```


4 Once continuous variable

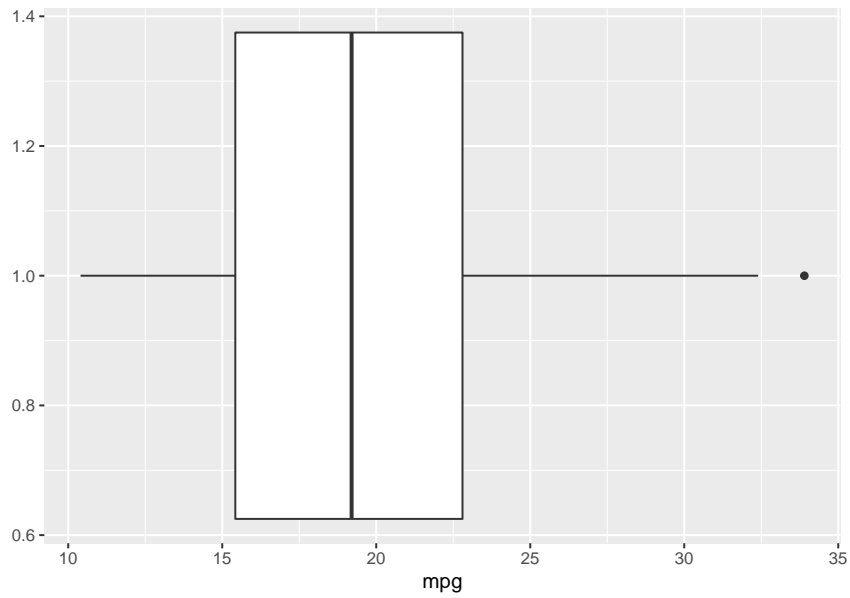
4.1 Histogram

```
data %>% ggplot(aes(x = mpg, y = stat(density))) + geom_histogram(aes(fill = stat(density)), binwidth =  
  scale_fill_distiller(type = "seq", direction = 1) +  
  geom_density(col = "brown", size = 1)
```



4.2 Boxplot

```
ggplot(data, aes(1,mpg)) +  
  geom_boxplot() +  
  coord_flip() +  
  xlab("")
```

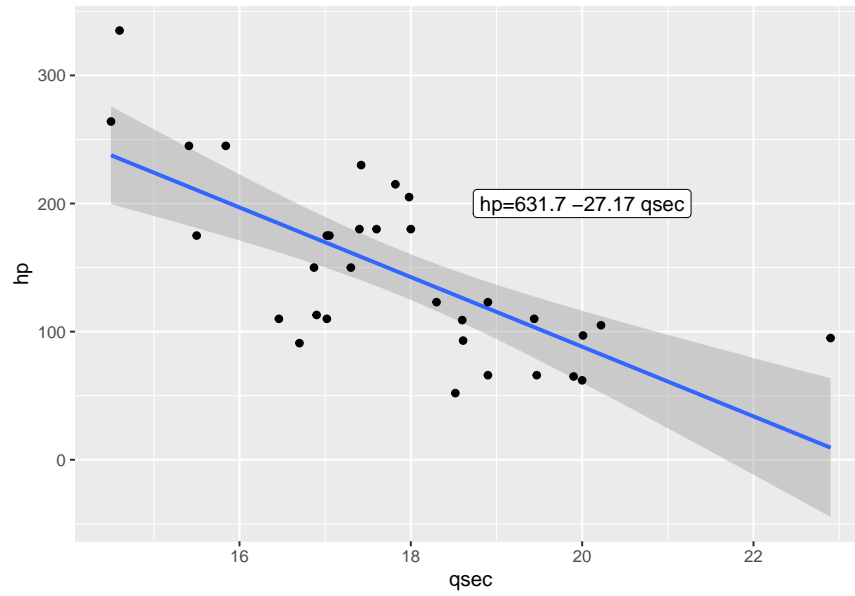


```
lm(hp ~ qsec, data = data)
```

```
##
## Call:
## lm(formula = hp ~ qsec, data = data)
##
## Coefficients:
## (Intercept)      qsec
##      631.70      -27.17
```

```
ggplot(data, aes(x = qsec, y = hp)) +
  geom_smooth(method = "lm") +
  geom_point() +
  annotate("label", 20, 200, label="hp=631.7 -27.17 qsec")
```

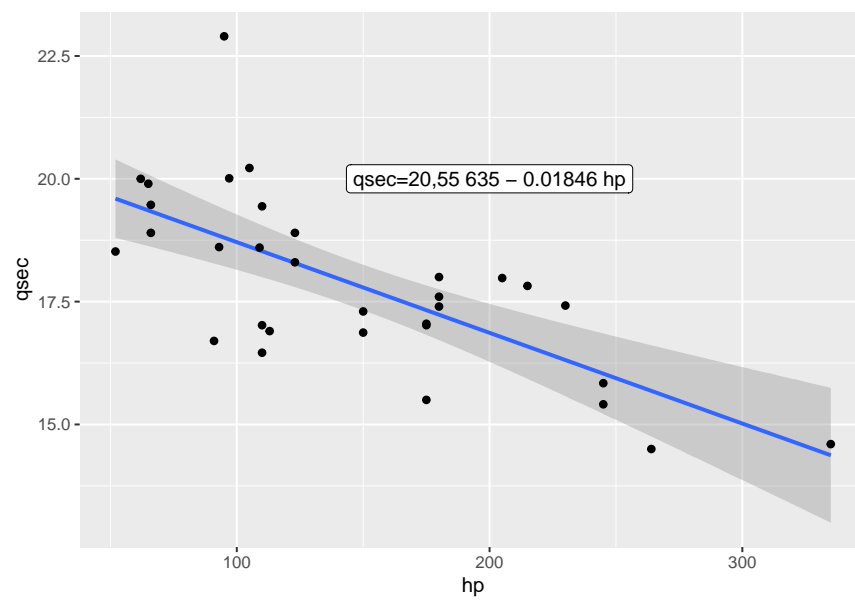
```
## 'geom_smooth()' using formula 'y ~ x'
```



4.3 Erklärende Variable sollten eig PS sein

```
ggplot(data, aes(x = hp, y = qsec)) +
  geom_smooth(method = "lm") +
  geom_point() +
  annotate("label", 200, 20, label="qsec=20,55 635 - 0.01846 hp")
```

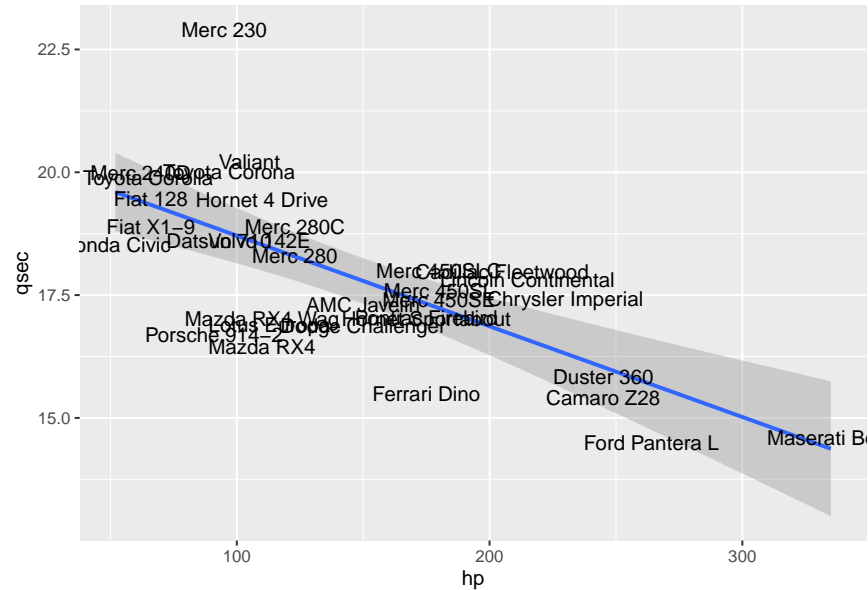
'geom_smooth()' using formula 'y ~ x'



4.4 Scatter plot using labels

```
data %>% rownames_to_column(var="name") %>%  
ggplot(aes(hp, qsec)) +  
  geom_smooth(method = "lm") +  
  geom_text(aes(label = name))
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

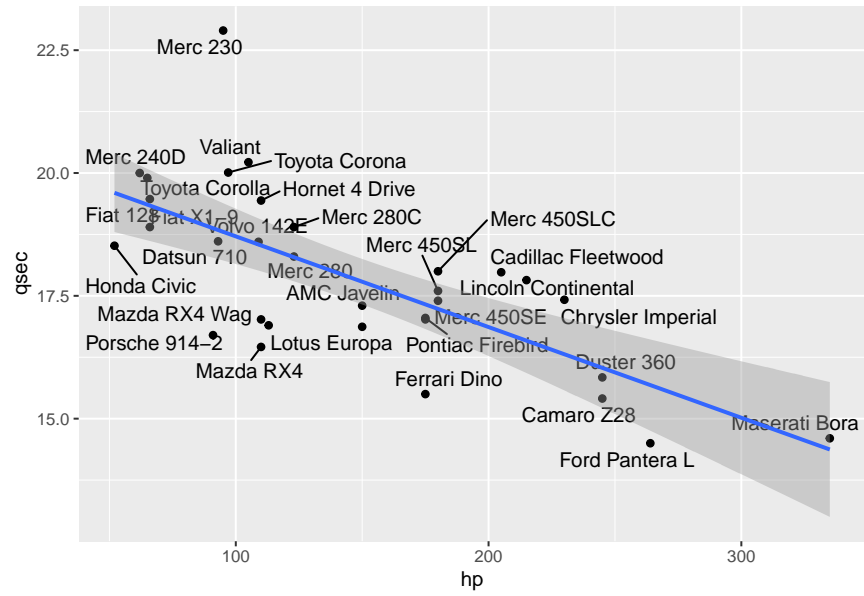


5 Using ggrepel package

```
library(ggrepel)  
data %>% rownames_to_column(var="name") %>% ggplot(aes(hp, qsec)) +  
  geom_point() +  
  geom_text_repel(aes(label = name)) +  
  geom_smooth(method = "lm")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

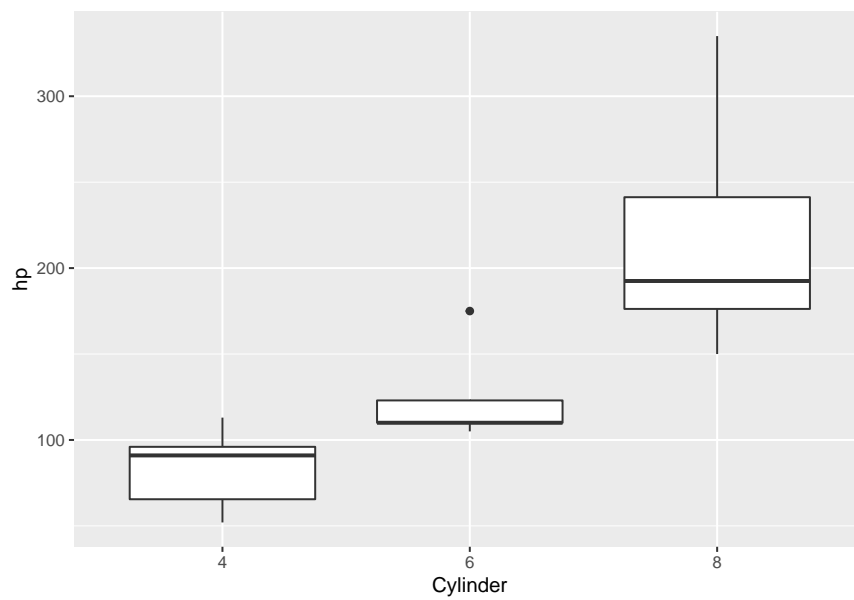
```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```



6 One categorical and one numeric variable

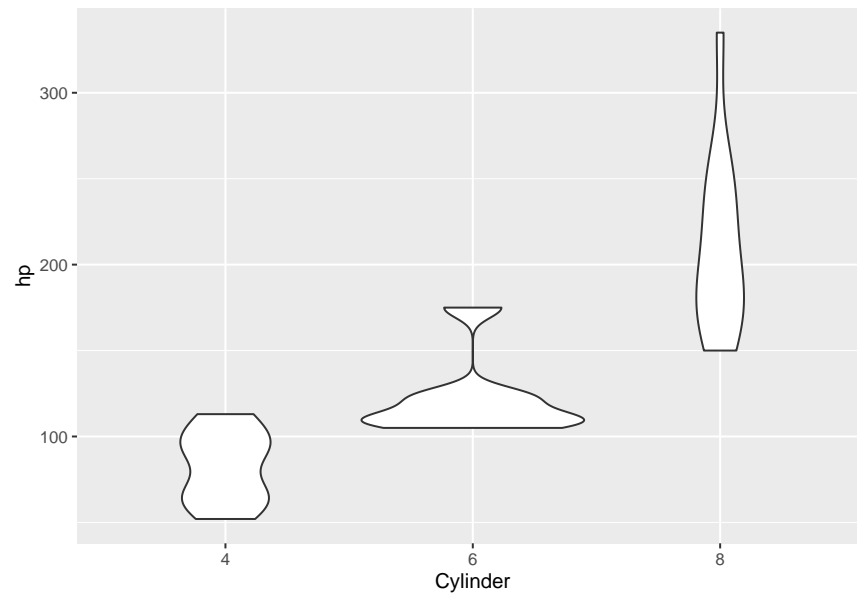
6.1 Boxplot

```
ggplot(data, aes(factor(cyl), hp)) +
  geom_boxplot() +
  xlab("Cylinder")
```



6.2 Violin plot

```
ggplot(data, aes(factor(cyl), hp)) +  
  xlab("Cylinder") +  
  geom_violin()
```

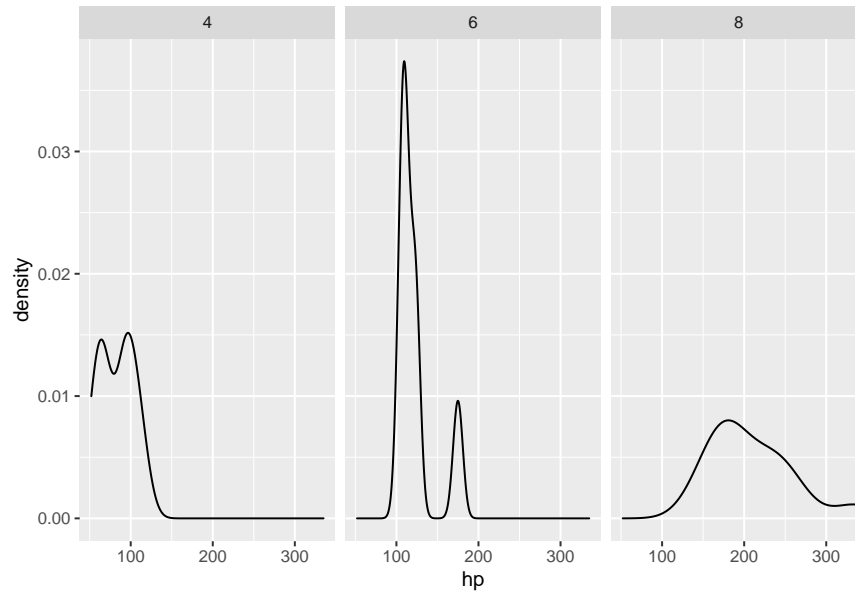


Interpretation:

- hp increases with number of cylinders
- very high horse power cars have all 8 cylinders

6.3 Facetted Plot

```
ggplot(data, aes(x = hp)) +  
  geom_density() +  
  facet_wrap(~cyl)
```



6.4 Exercise 2

Visualize calories, given manufacturer

```
head(UScereal, 1)
```

```
##           mfr calories protein      fat  sodium   fibre   carbo  sugars
## 100% Bran   N  212.1212 12.12121  3.030303 393.9394 30.30303 15.15152 18.18182
##           shelf potassium vitamins
## 100% Bran     3   848.4849 enriched
```

```
UScereal %>% ggplot(aes(x = calories, fill=mfr)) + geom_boxplot()
```

