# R5: Visualising variables

# Data visualisation with ggplot2 in tidyverse

### Andreas Reschreiter

# Contents

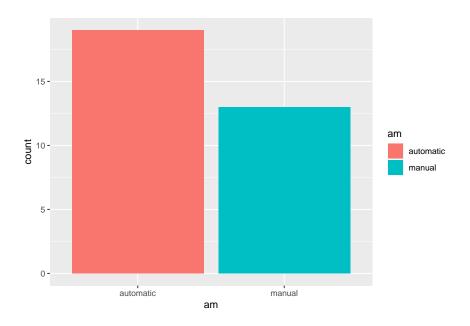
1	One categorical variable		2
	1.1	Classic bar chart	2
	1.2	Stacked bar chart	2
	1.3	Pie chart (deprecated)	3
2	Two categorical variables		4
	2.1	Stacked bar charts	4
	2.2	Stacked bar charts with proportions	4
	2.3	Grouped bar chart	5
	2.4	Grouped bar chart with proportions (better use stacked bar chart instead!)	5
3	One	e continous variable	6
	3.1	Histogram (with density plot):	6
	3.2	Boxplot:	6
4	Two continuous variables		
	4.1	Standard scatter plot with regression line:	7
	4.2	Scatter plot Using text labels:	8
5	One continuous, one discrete		
	5.1	Parallel boxplots:	9
	5.2	Facetted density plots:	10
6	Special Topics		
	6.1	Spine plots / Mosaic plots	11
	6.2	Time series	13

Packages and data used in this notebook:

# 1 One categorical variable

#### 1.1 Classic bar chart

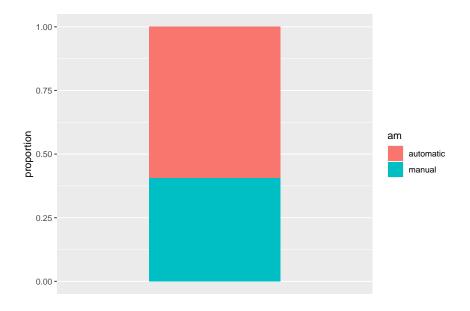
```
ggplot(cars, aes(x = am, fill = am)) + geom_bar()
```



Note: geom\_bar() implicitely computes the counts and maps them to y!

#### 1.2 Stacked bar chart

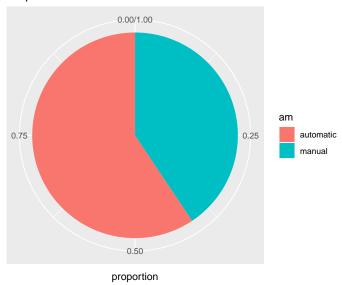
```
ggplot(cars, aes(x = "x", fill = am)) +
geom_bar(position = "fill", width = 0.5) +
ylab("proportion") + xlab("") +
scale_x_discrete(breaks = NULL)
```



## 1.3 Pie chart (deprecated)

```
ggplot(cars, aes(x = 1, fill = am)) +
  geom_bar(position = "fill") +
  ylab("proportion") + coord_polar("y") +
  xlab("") + scale_x_continuous(breaks = NULL) +
  ggtitle("Proportion of automatic and manual transmission")
```

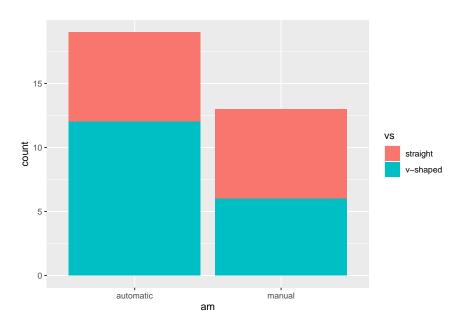
#### Proportion of automatic and manual transmission



# 2 Two categorical variables

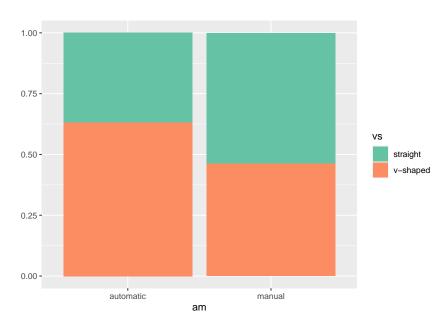
#### 2.1 Stacked bar charts

```
ggplot(cars, aes(x = am, fill = vs)) + geom_bar()
```



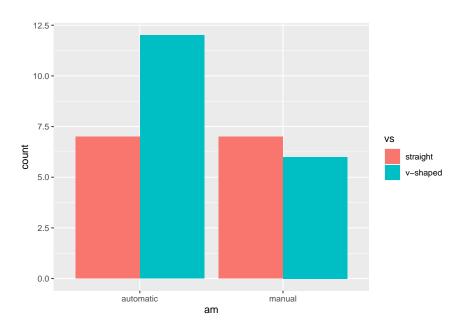
## 2.2 Stacked bar charts with proportions

```
ggplot(cars, aes(x = am, fill = vs)) + geom_bar(position = "fill") +
    ylab("") +
    scale_fill_brewer(type = "qual", palette = 7)
```



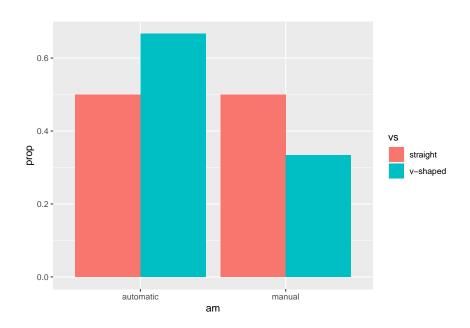
### 2.3 Grouped bar chart

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



## 2.4 Grouped bar chart with proportions (better use stacked bar chart instead!)

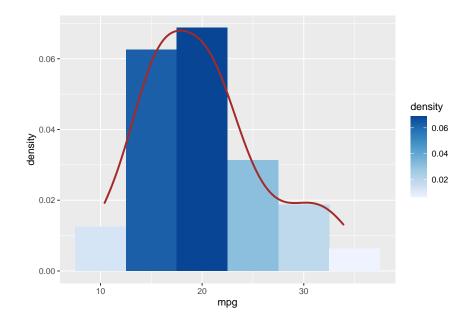
```
ggplot(cars, aes(x = am, y = stat(prop), fill = vs, group = vs)) +
  geom_bar(position = "dodge")
```



#### 3 One continous variable

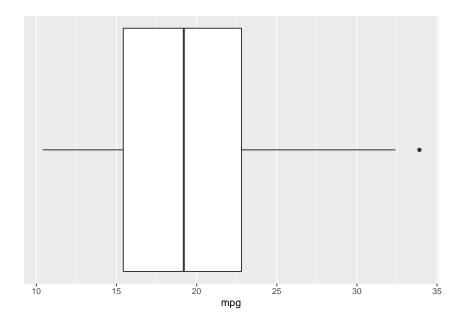
### 3.1 Histogram (with density plot):

```
ggplot(cars, aes(x = mpg, y = stat(density))) + ## uses density instead of counts
# geom_histogram(binwidth = 5) +
geom_histogram(aes(fill = stat(x)), binwidth = 5) +
geom_histogram(aes(fill = stat(density)), binwidth = 5) +
scale_fill_distiller(type = "seq", direction = 1) +
geom_density(col = "brown", size = 1)
```



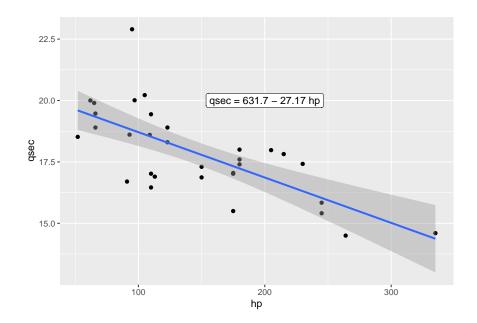
#### 3.2 Boxplot:

```
ggplot(cars, aes(1, mpg)) + ## x mapped to 1 since there are no groups
geom_boxplot() +
coord_flip() +
    ## remove labels and ticks on the x scale
xlab("") + scale_x_continuous(breaks = NULL)
```



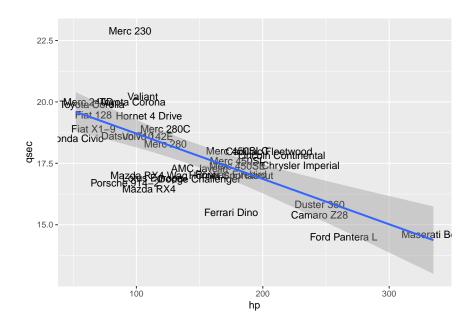
## 4 Two continuous variables

### 4.1 Standard scatter plot with regression line:



## 4.2 Scatter plot Using text labels:

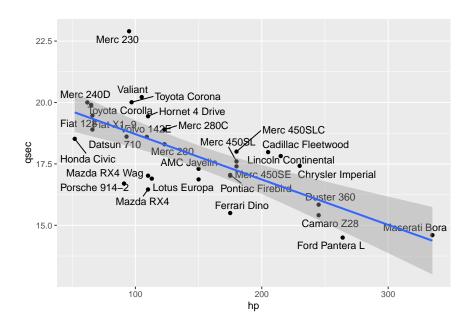
```
ggplot(cars, aes(hp, qsec)) +
  geom_text(aes(label = names)) +
  geom_smooth(method = "lm")
```



Better solution using ggrepel package:

```
library("ggrepel")
ggplot(cars, aes(hp, qsec)) +
  geom_point() +
```

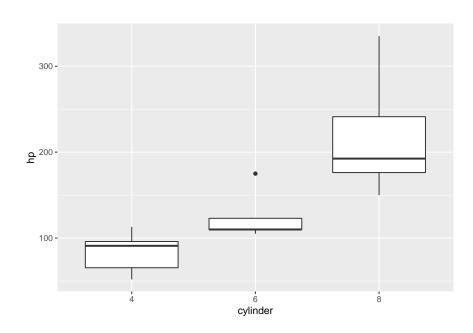
```
geom_text_repel(aes(label = names)) +
geom_smooth(method = "lm")
```

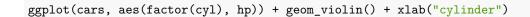


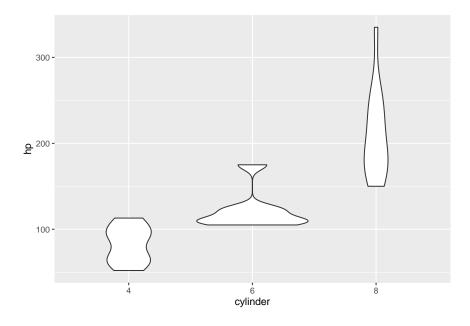
# 5 One continuous, one discrete

### 5.1 Parallel boxplots:

```
ggplot(cars, aes(factor(cyl), hp)) + geom_boxplot() + xlab("cylinder")
```

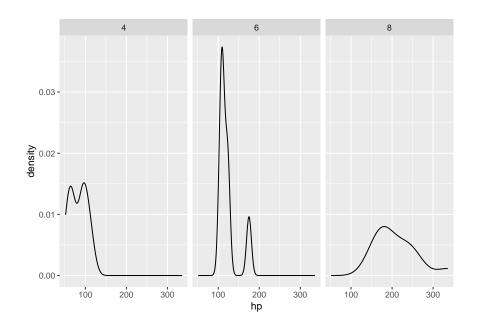






# 5.2 Facetted density plots:

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



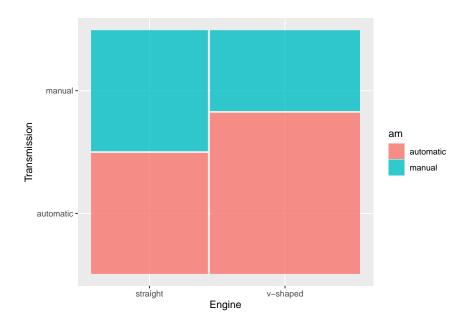
## 6 Special Topics

### 6.1 Spine plots / Mosaic plots

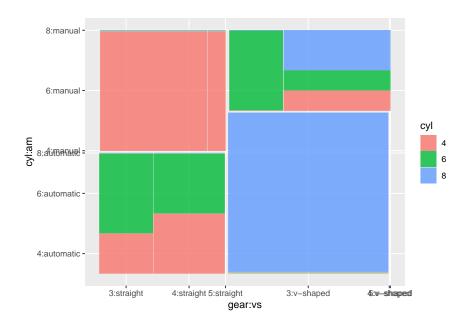
#### 6.1.1 Cars data

A spine plot is a special case of a mosaic plot:

```
library("ggmosaic")
ggplot(cars) + geom_mosaic(aes(product(vs), fill = am)) +
    xlab("Engine") + ylab("Transmission") # + scale_y_continuous()
```



Mosaic plots can handle more than two variables:

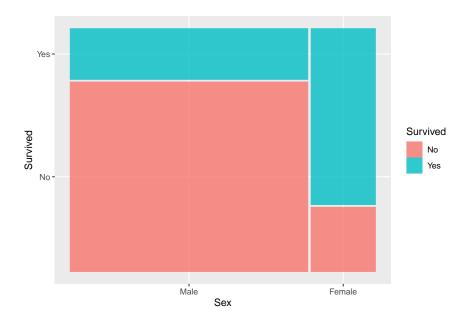


#### 6.1.2 Titanic Data

```
data(titanic)
```

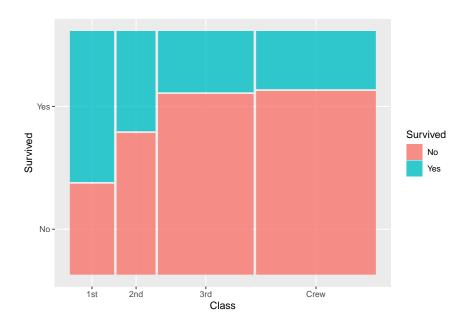
Plot for Survival and Sex

```
ggplot(data = titanic) +
geom_mosaic(aes(x = product(Survived, Sex), fill = Survived))
```



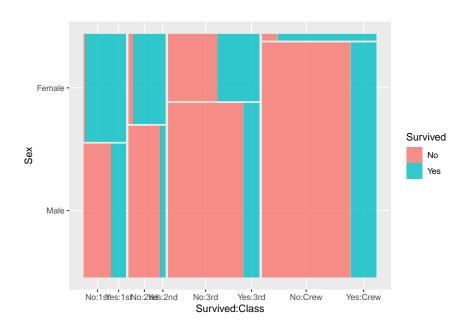
Plot for Survival and Class

```
ggplot(data = titanic) +
geom_mosaic(aes(x = product(Survived, Class), fill = Survived))
```



Plot for Survival, Sex and Class

```
ggplot(data = titanic) +
geom_mosaic(aes(x = product(Survived, Sex, Class), fill = Survived))
```



#### 6.2 Time series

```
library(zoo)
as_tibble(Seatbelts) %>%
  mutate(date = as.Date(Seatbelts)) %>%
  gather(key = "Position", value = "Deaths", front, rear) %>%

ggplot(aes(date, Deaths, col = Position)) +
  geom_line() +
  scale_x_date(breaks = "3 years") +
  geom_vline(xintercept = as.Date("1983-01-01"), col = "blue")
```

