Ggplot2 tutorial - Command lines

Ariane Ducellier

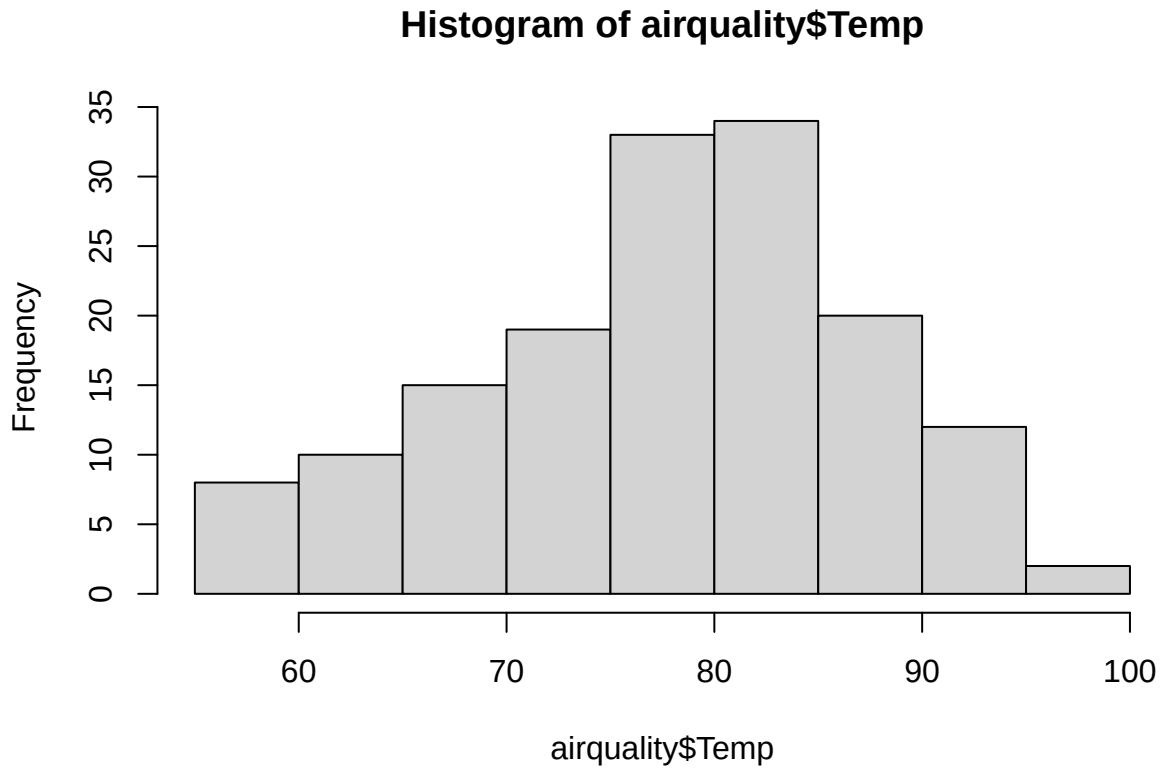
7/18/2023

```
library(ggplot2)
library(Lock5Data)
```

Part 1 - Basic plotting in ggplot2

Histograms

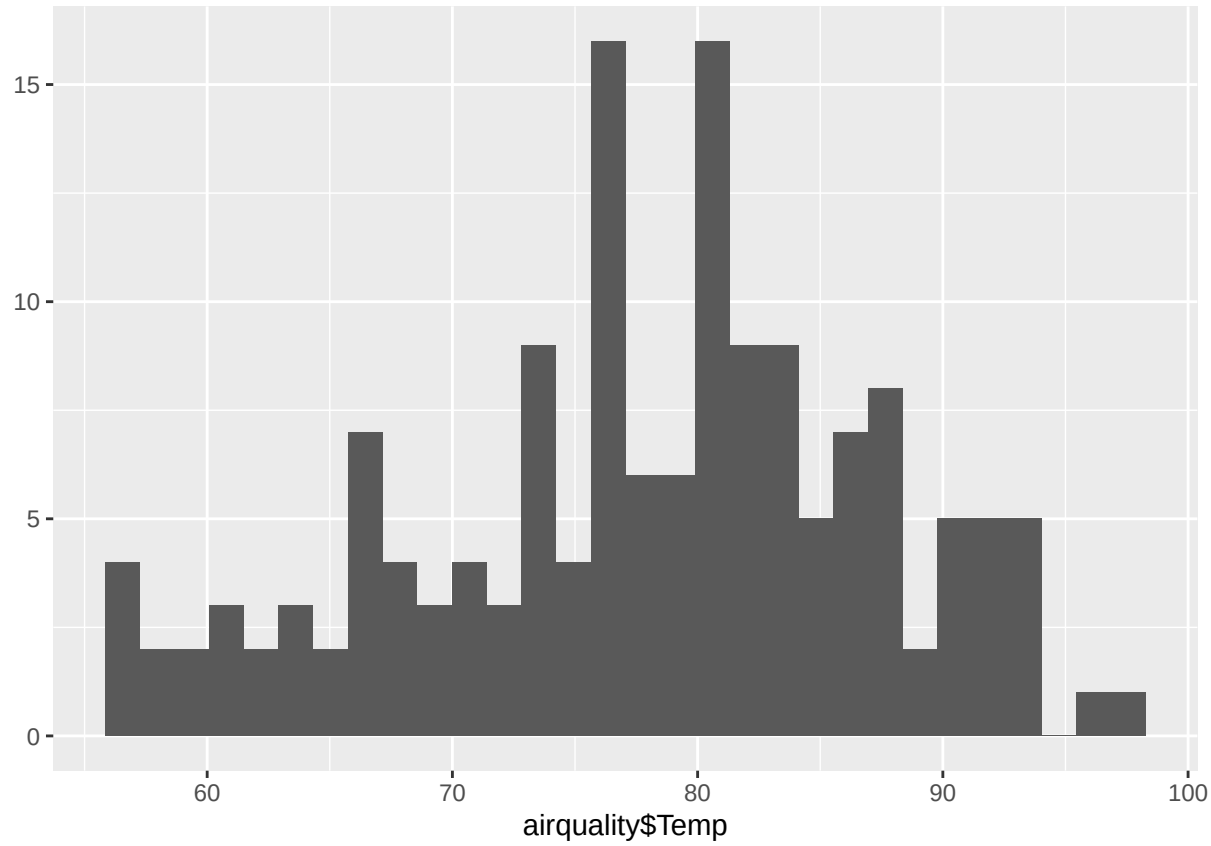
```
hist(airquality$Temp)
```



```
qplot(airquality$Temp)
```

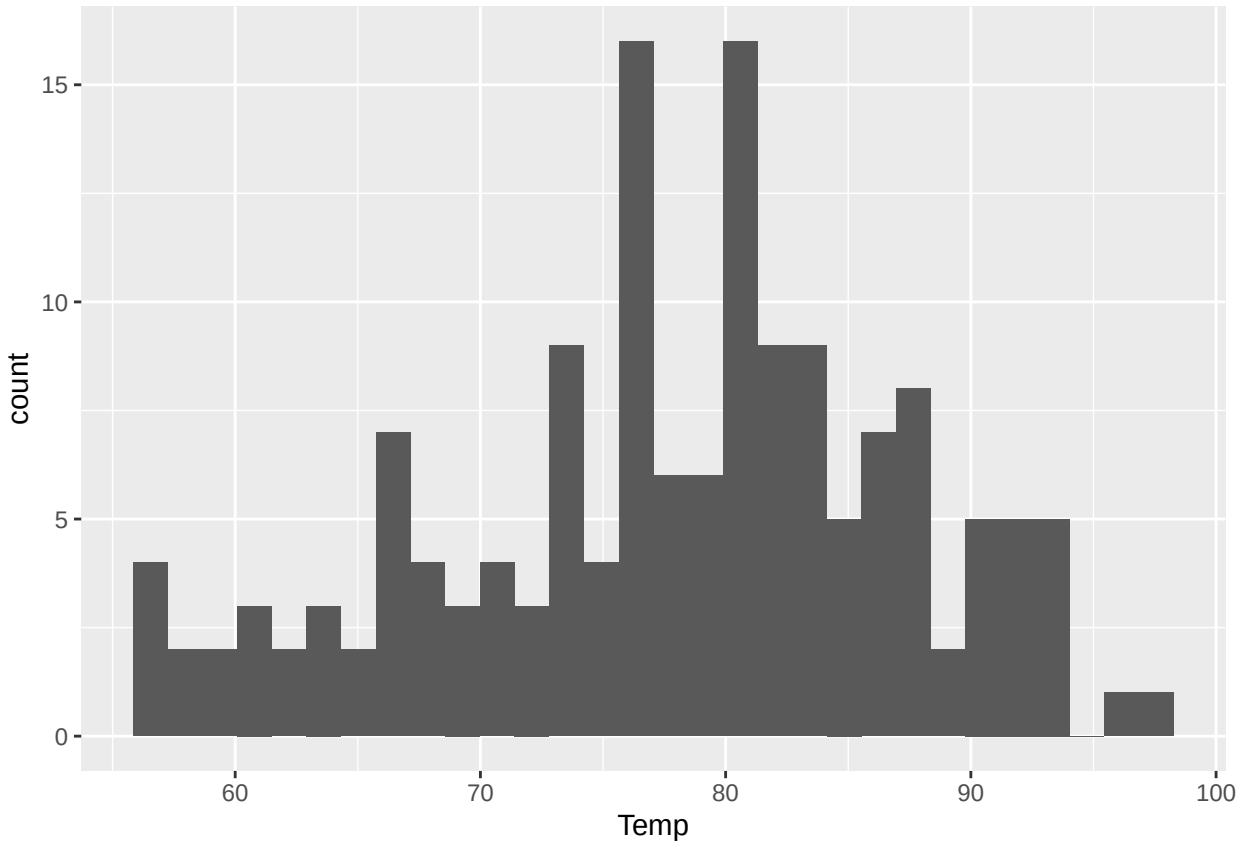
```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```

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



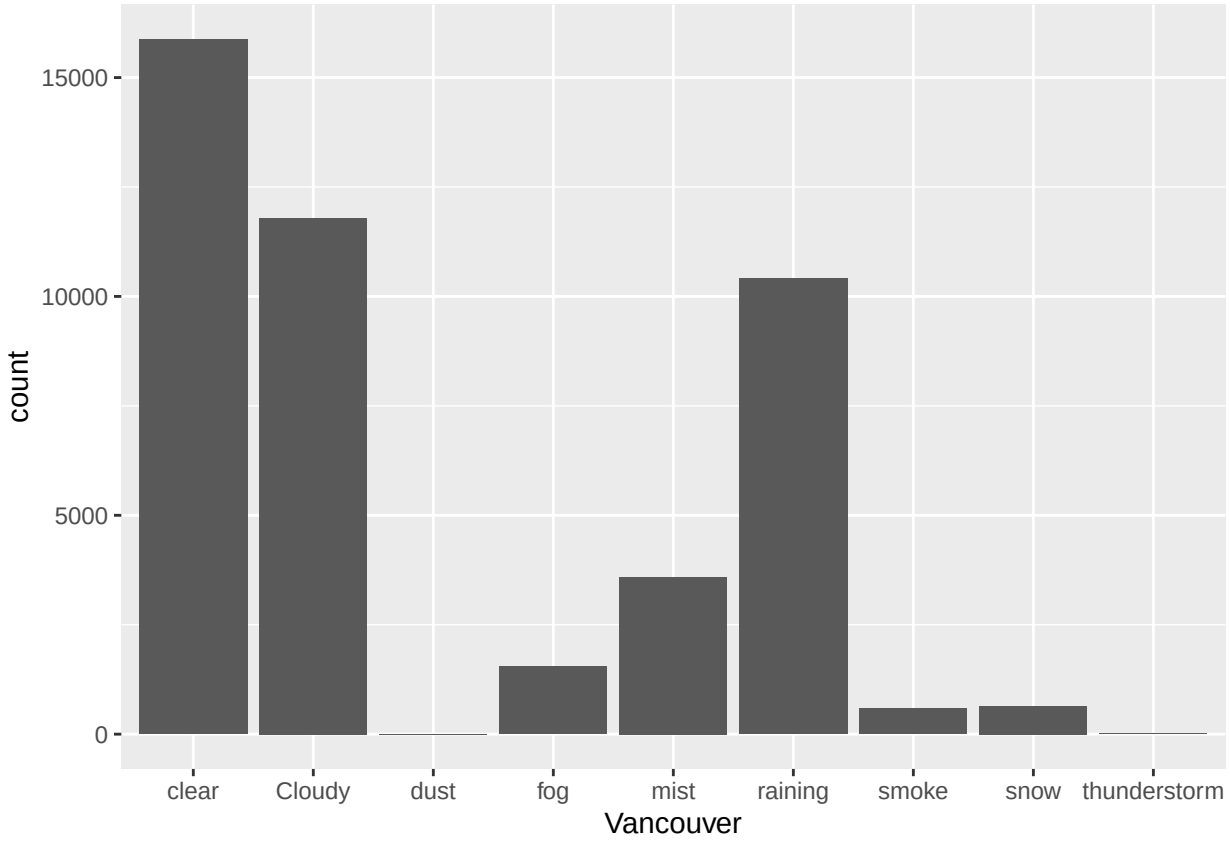

```
ggplot(airquality, aes(x=Temp)) + geom_histogram()
```

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



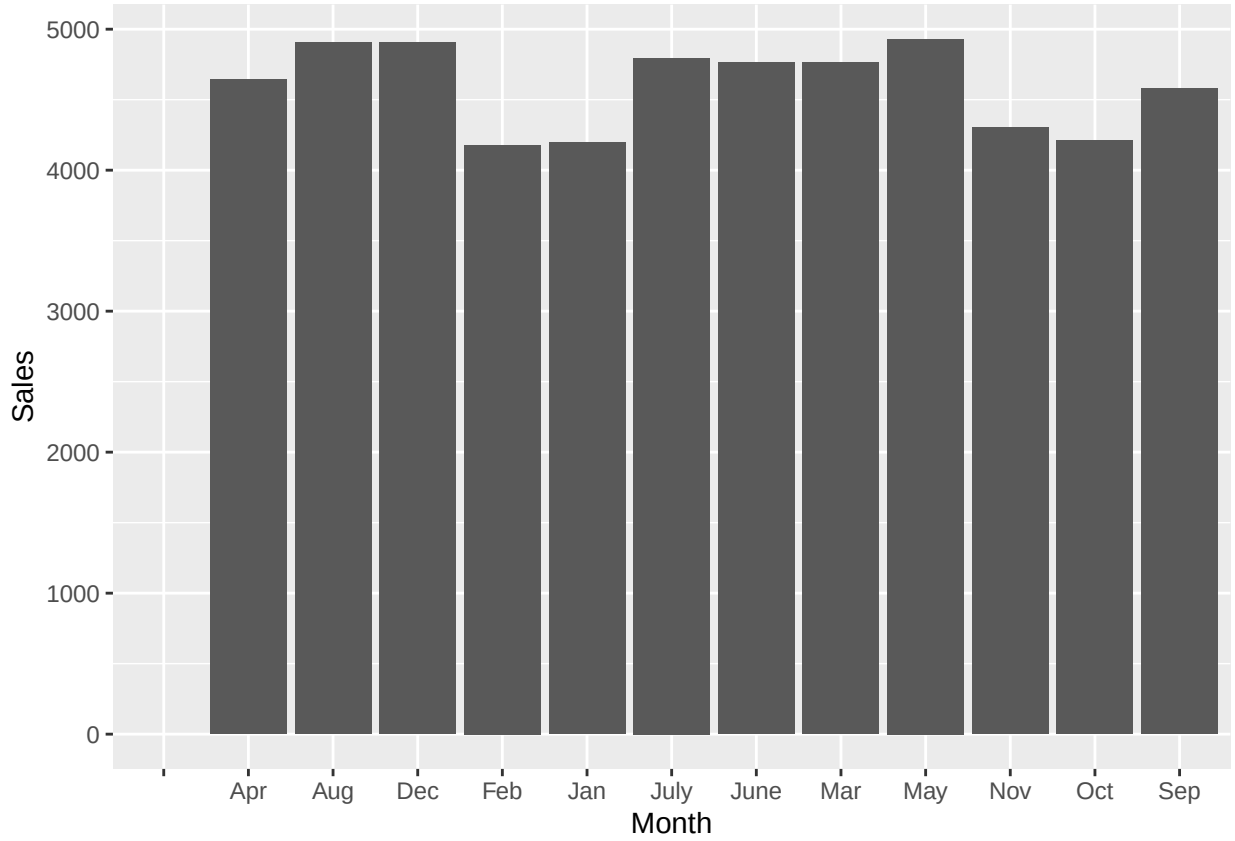
Bar plots

```
df_desc <- read.csv("../data/historical-hourly-weather-data/weather_description.csv")
ggplot(df_desc, aes(x=Vancouver)) + geom_bar()
```




```
ggplot(RetailSales, aes(x=Month, y=Sales)) + geom_bar(stat="identity")
```

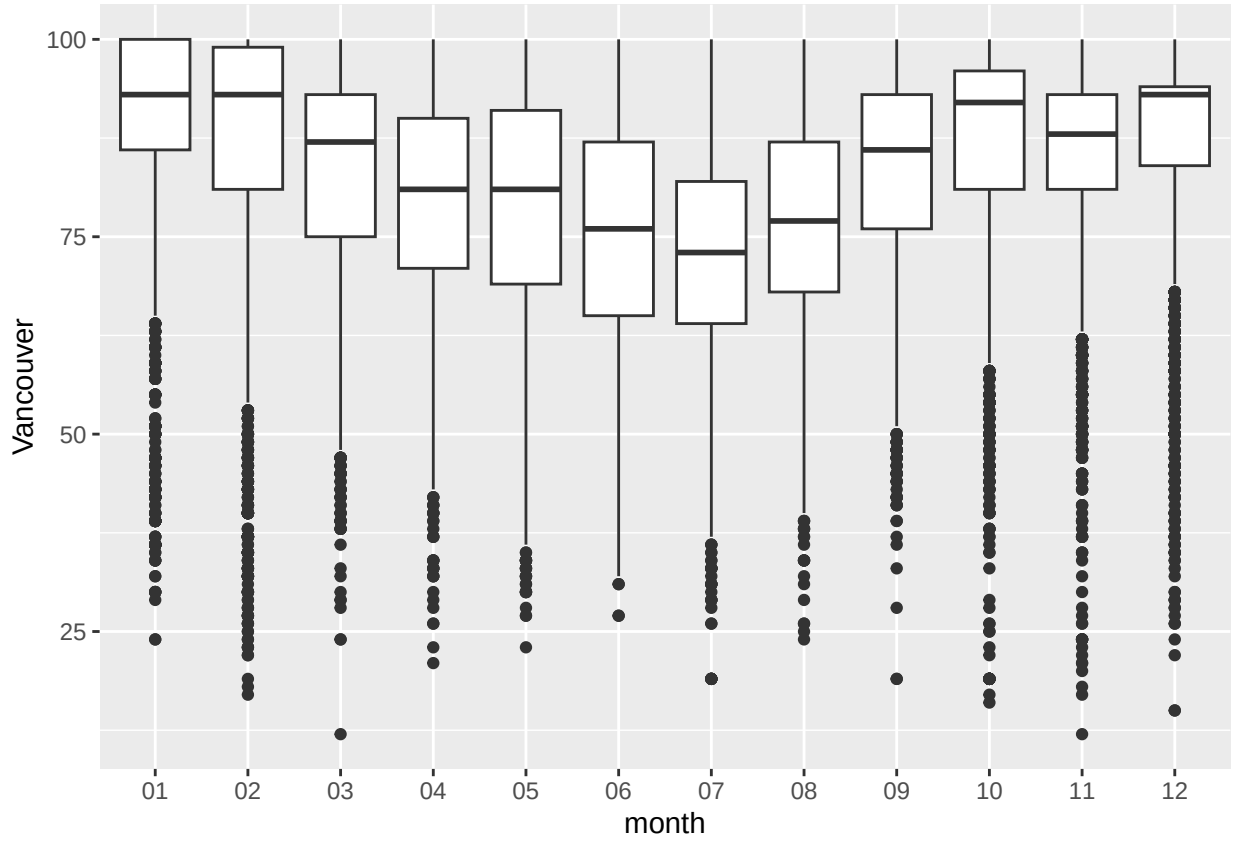
```
## Warning: Removed 15 rows containing missing values (‘position_stack()‘).
```



Box plots

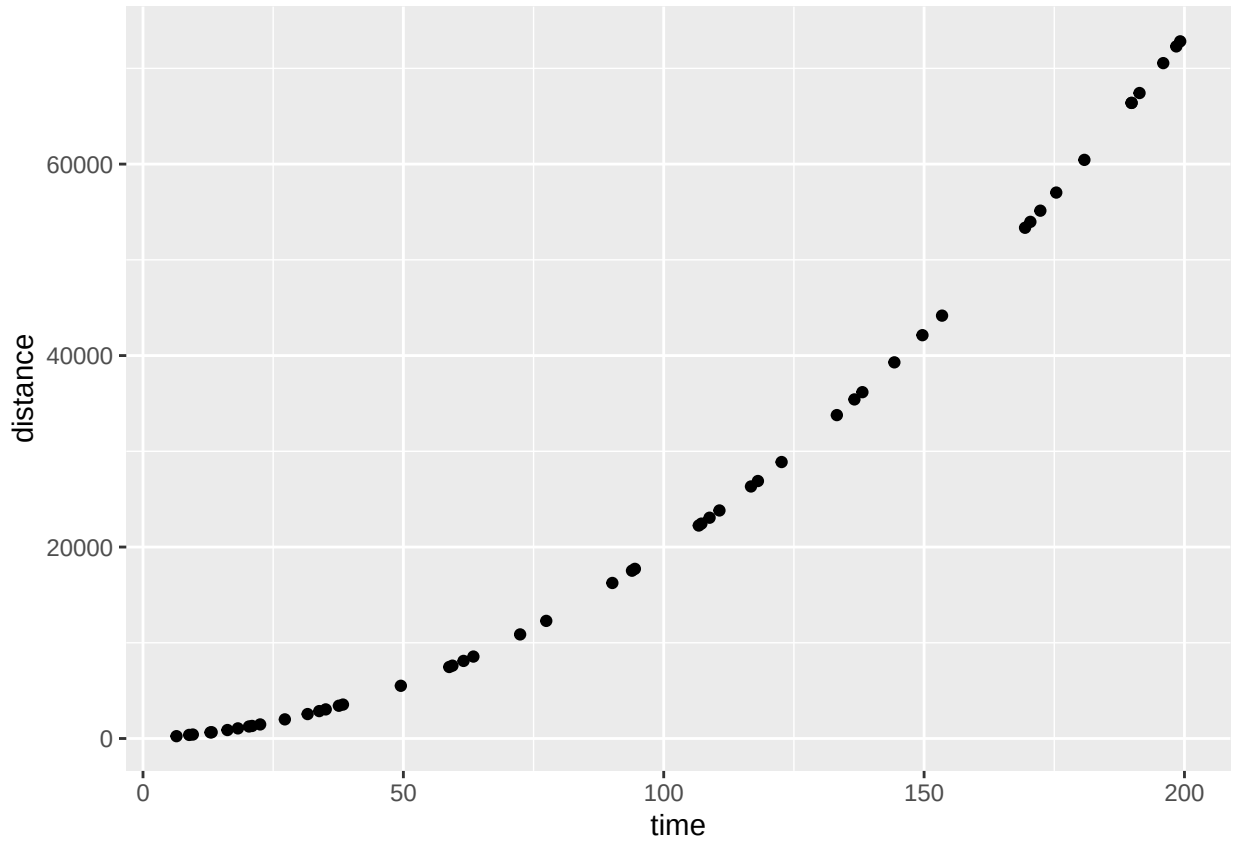
```
df_hum <- read.csv("../data/historical-hourly-weather-data/humidity.csv")
df_hum$datetime <- as.character(df_hum$datetime)
df_hum$month <- substr(df_hum$datetime, 6, 7)
ggplot(df_hum, aes(x=month, y=Vancouver)) + geom_boxplot()
```

```
## Warning: Removed 1826 rows containing non-finite values (‘stat_boxplot()‘).
```

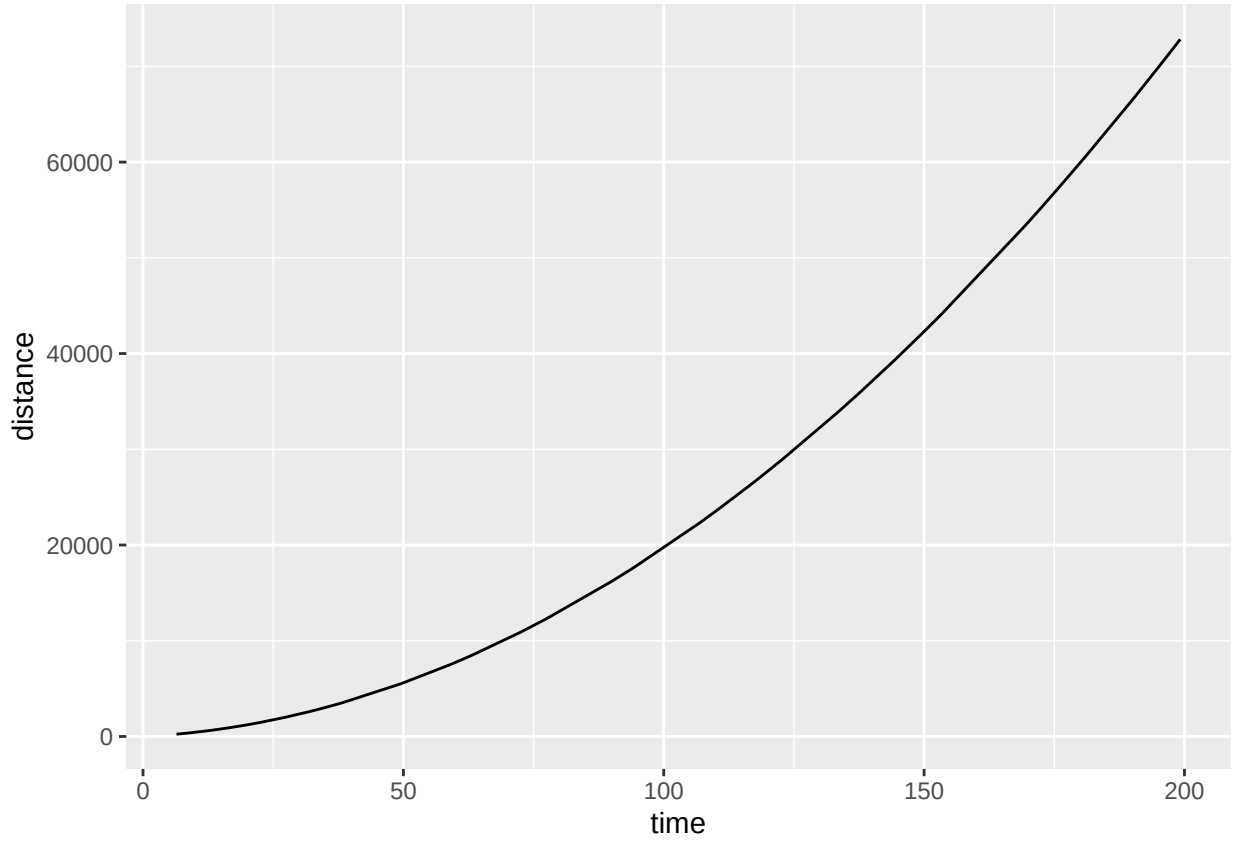


Scatter plots and line plots

```
a = 3.4
v0 = 27
time <- runif(50, min=0, max=200)
distance <- sapply(time, function(x) v0 * x + 0.5 * a * x^2)
df <- data.frame(time,distance)
ggplot(df, aes(x=time, y=distance)) + geom_point()
```




```
ggplot(df, aes(x=time, y=distance)) + geom_line()
```

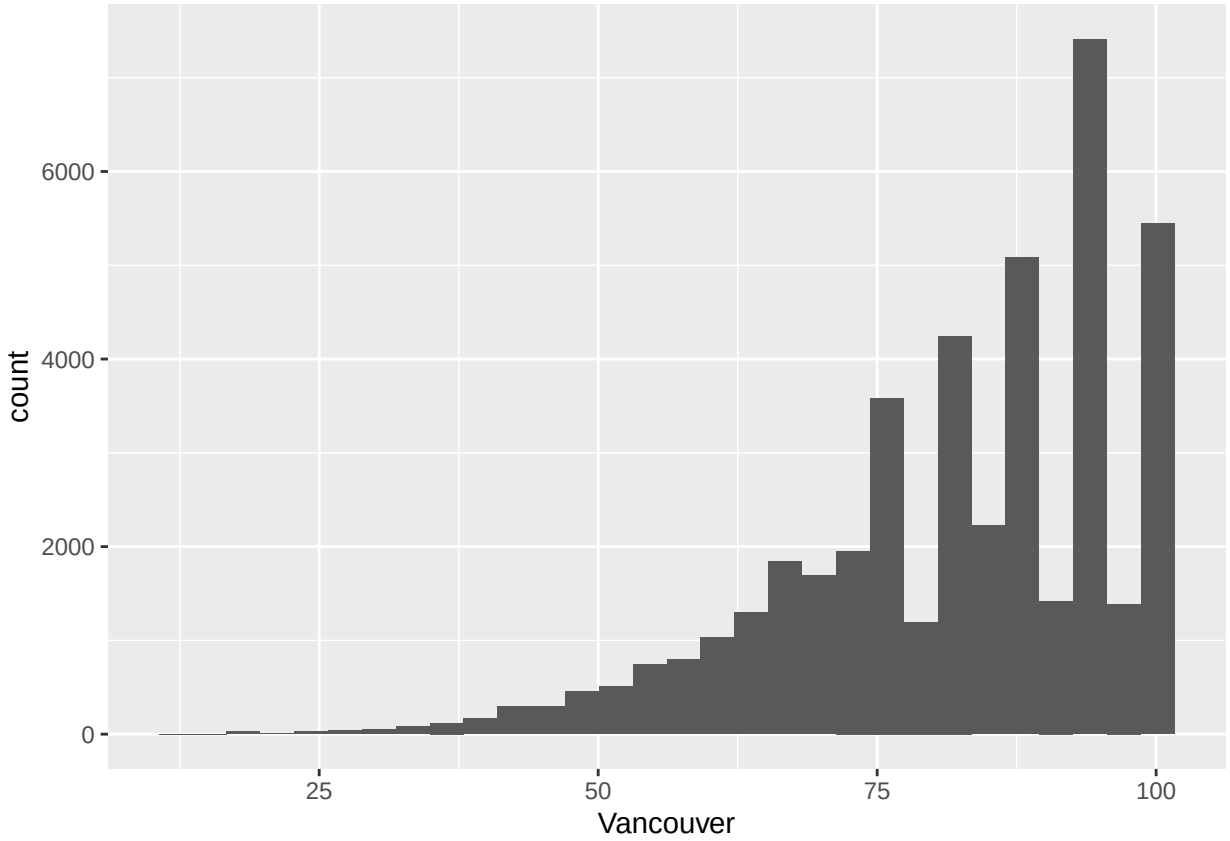


Changing histogram defaults and adding aesthetics

```
df_hum <- read.csv("../data/historical-hourly-weather-data/humidity.csv")
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram()
```

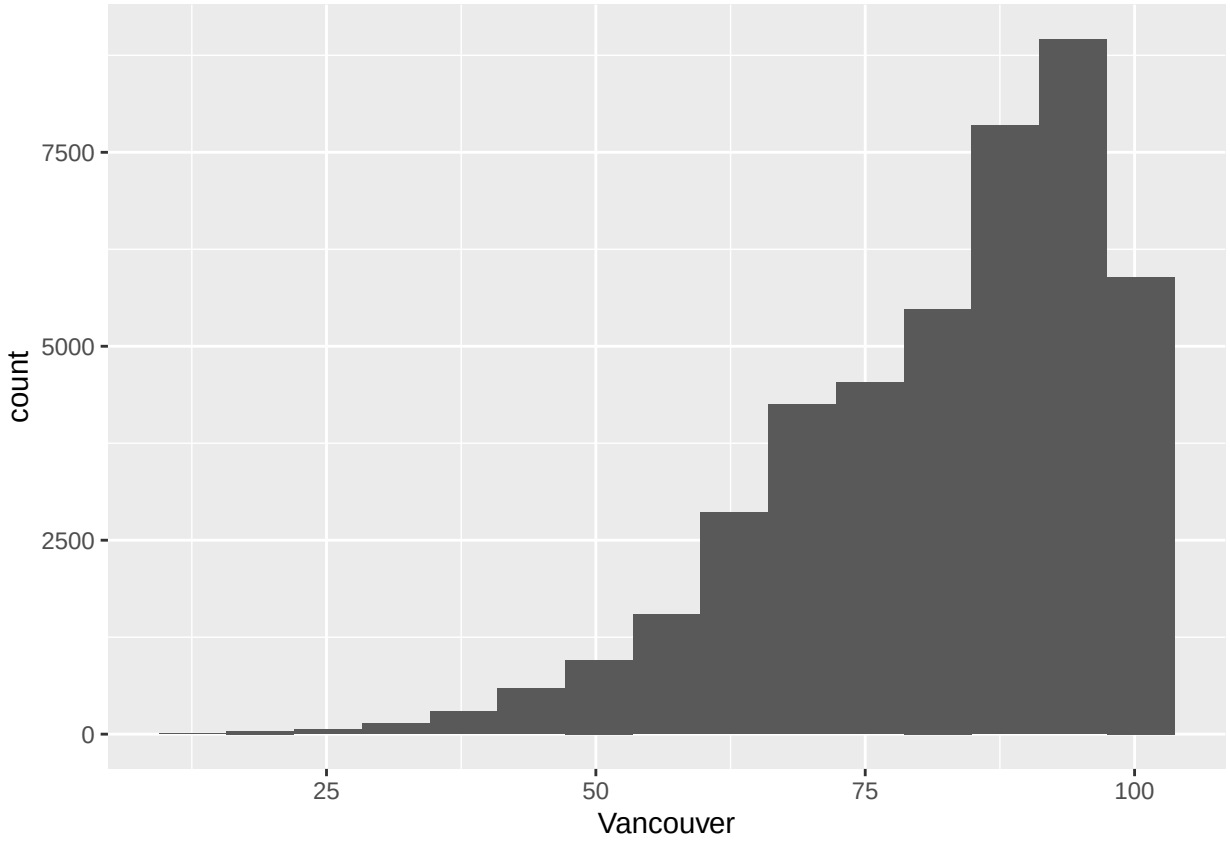
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1826 rows containing non-finite values (`stat_bin()`).
```



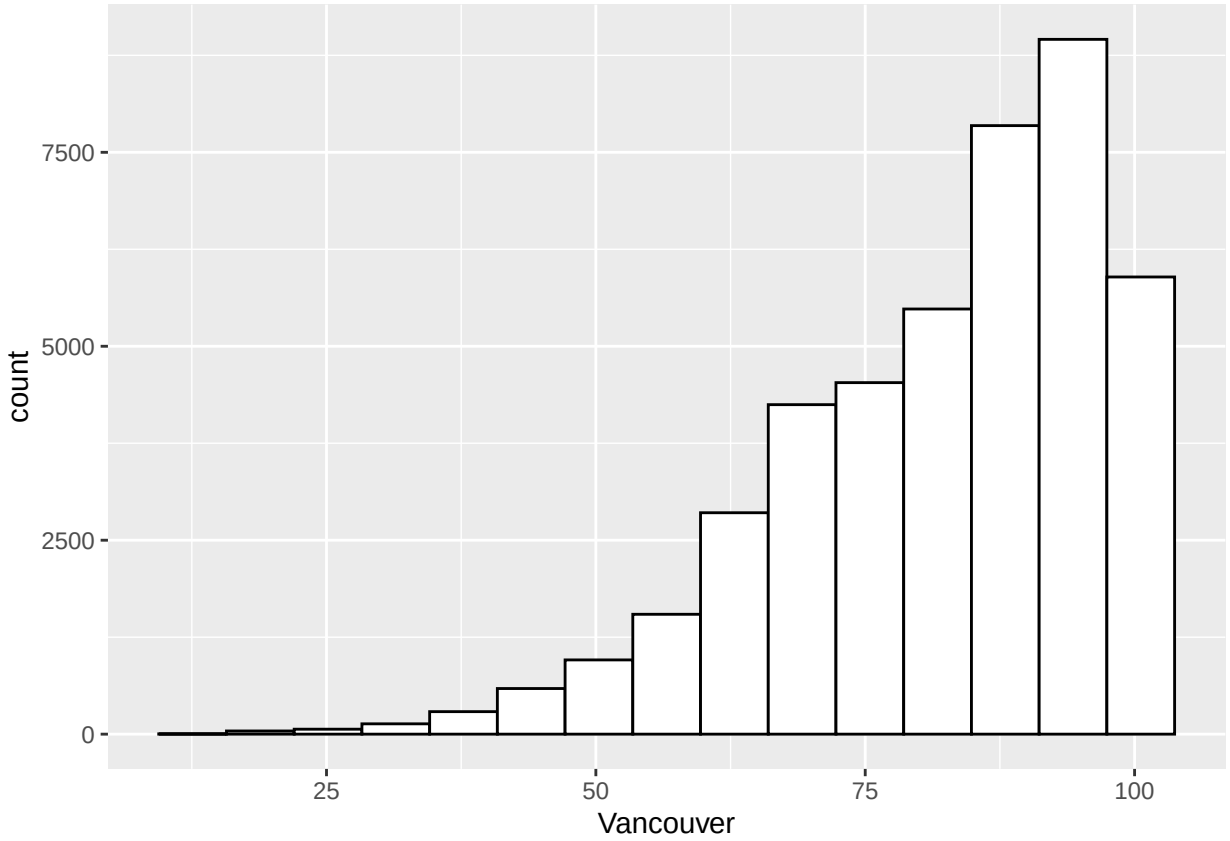

```
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram(bins=15)
```

```
## Warning: Removed 1826 rows containing non-finite values (‘stat_bin()‘).
```



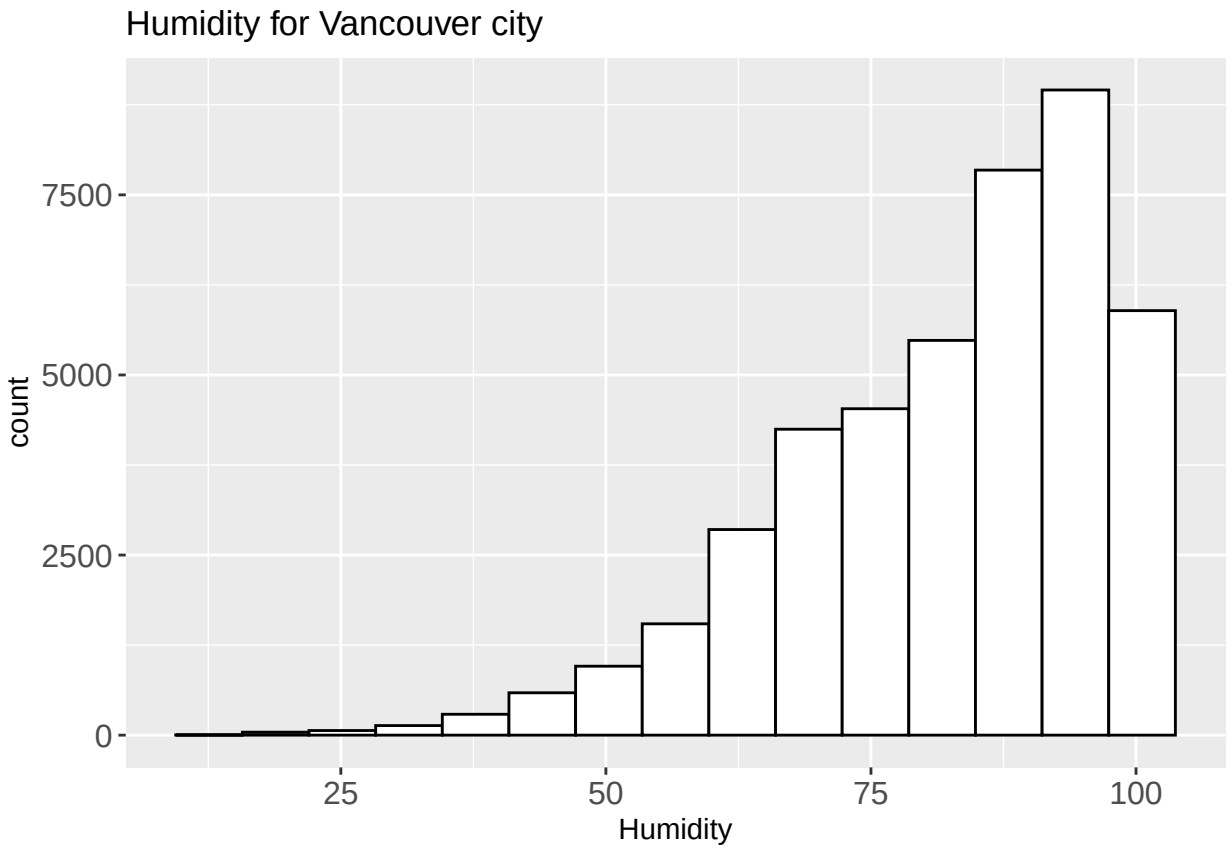

```
ggplot(df_hum, aes(x=Vancouver)) + geom_histogram(bins=15, fill="white", color=1)
```

```
## Warning: Removed 1826 rows containing non-finite values (‘stat_bin()‘).
```




```
ggplot(df_hum, aes(x=Vancouver)) +
geom_histogram(bins=15, fill="white", color=1) +
ggtitle("Humidity for Vancouver city") +
xlab("Humidity") +
theme(axis.text.x=element_text(size=12), axis.text.y=element_text(size=12))
```

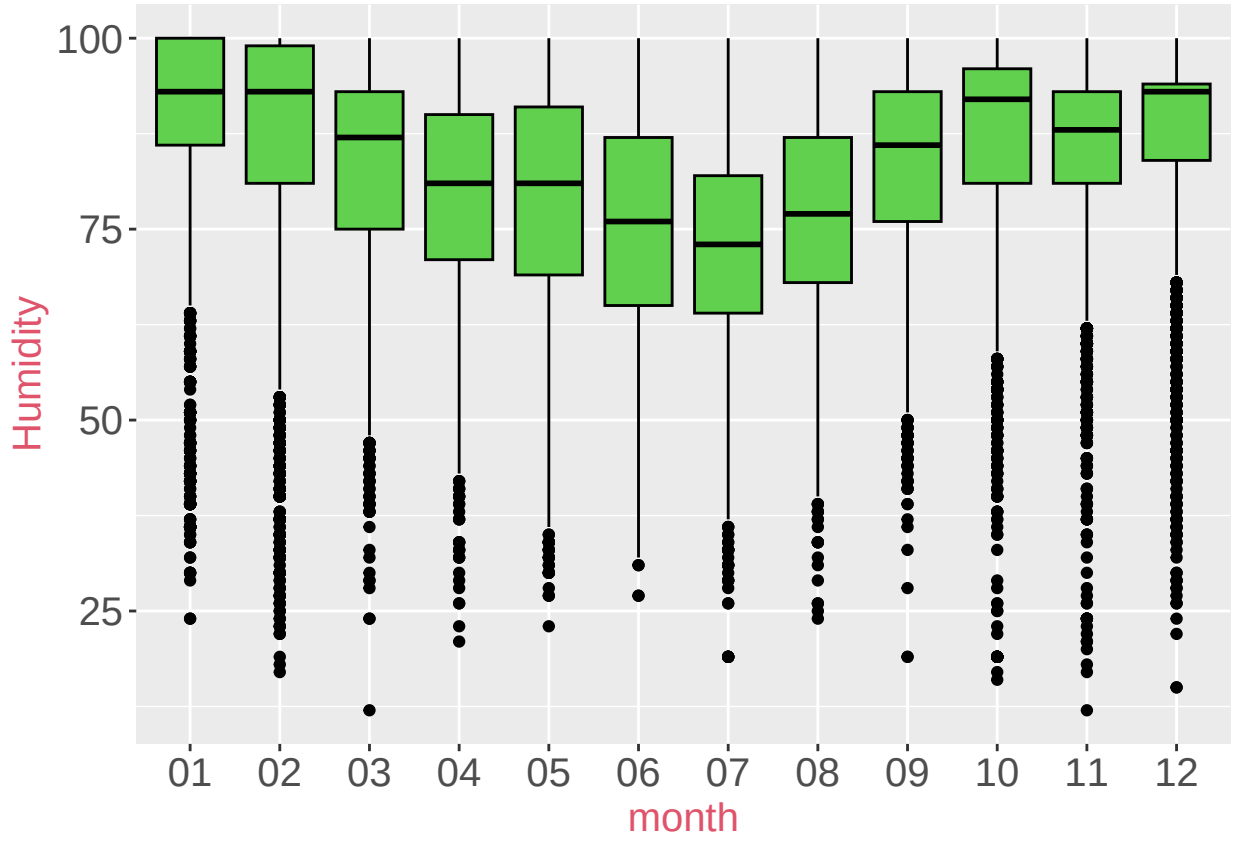
```
## Warning: Removed 1826 rows containing non-finite values (‘stat_bin()‘).
```



Changing boxplot defaults and adding aesthetics

```
df_hum <- read.csv("../data/historical-hourly-weather-data/humidity.csv")
df_hum$datetime <- as.character(df_hum$datetime)
df_hum$month <- substr(df_hum$datetime, 6, 7)
ggplot(df_hum, aes(x=month, y=Vancouver)) +
geom_boxplot(color=1, fill=3) +
ylab("Humidity") +
theme(axis.text.x=element_text(size=15),
axis.text.y=element_text(size=15),
axis.title.x=element_text(size=15, color=2),
axis.title.y=element_text(size=15, color=2))
```

```
## Warning: Removed 1826 rows containing non-finite values (‘stat_boxplot()‘).
```



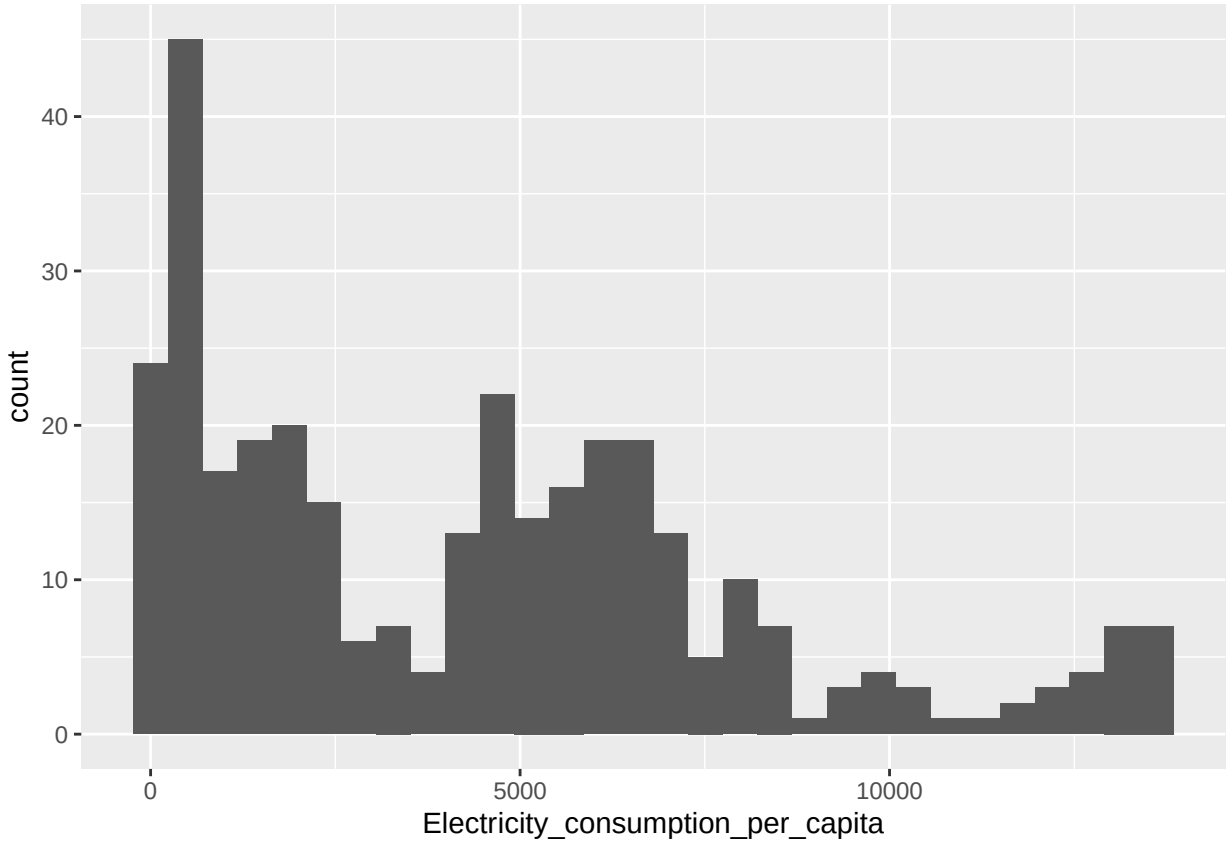
Part 2 - Grammar of graphics and visual components

Layers

```
df <- read.csv("../data/gapminder-data.csv")
p1 <- ggplot(df, aes(x=Electricity_consumption_per_capita))
p2 <- p1 + geom_histogram()
p3 <- p1 + geom_histogram(bins=15)
p4 <- p3 + xlab("Electricity consumption per capita")
p2
```

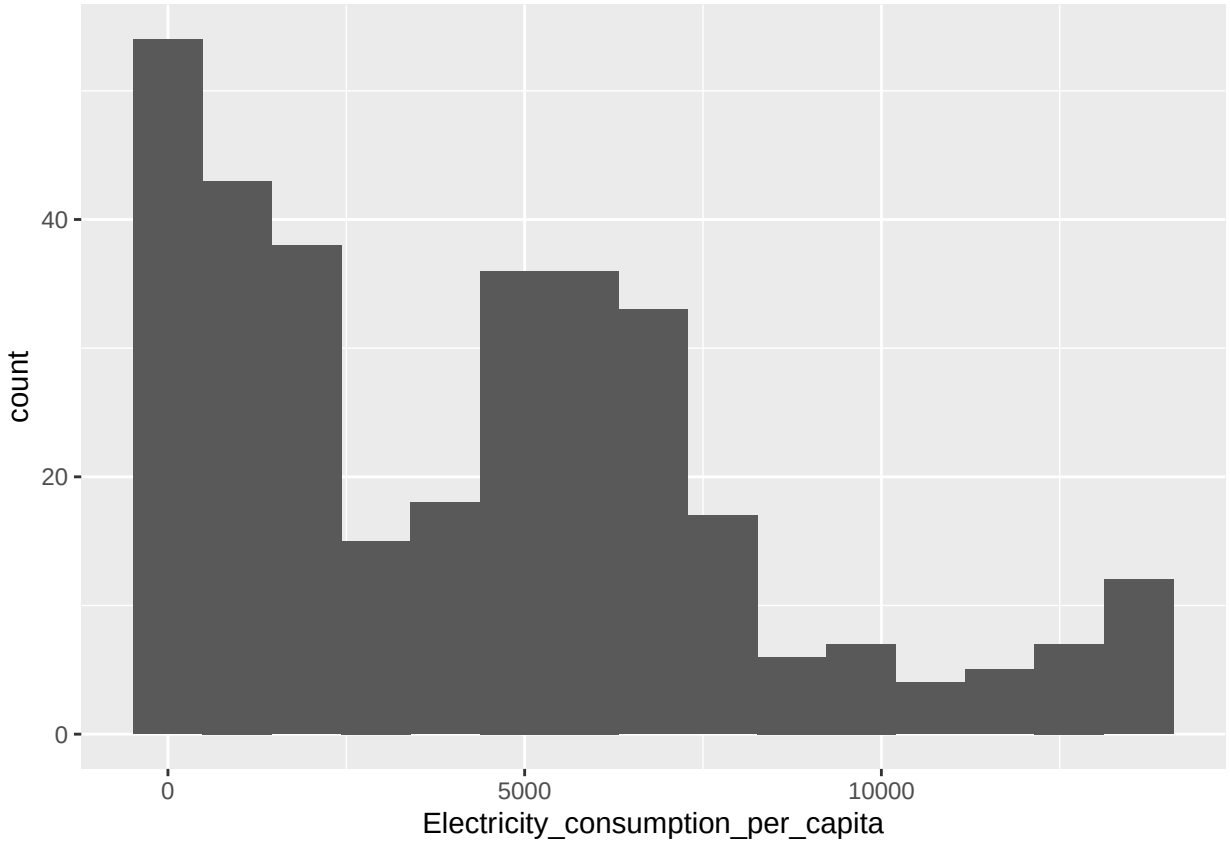
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1181 rows containing non-finite values (`stat_bin()`).
```



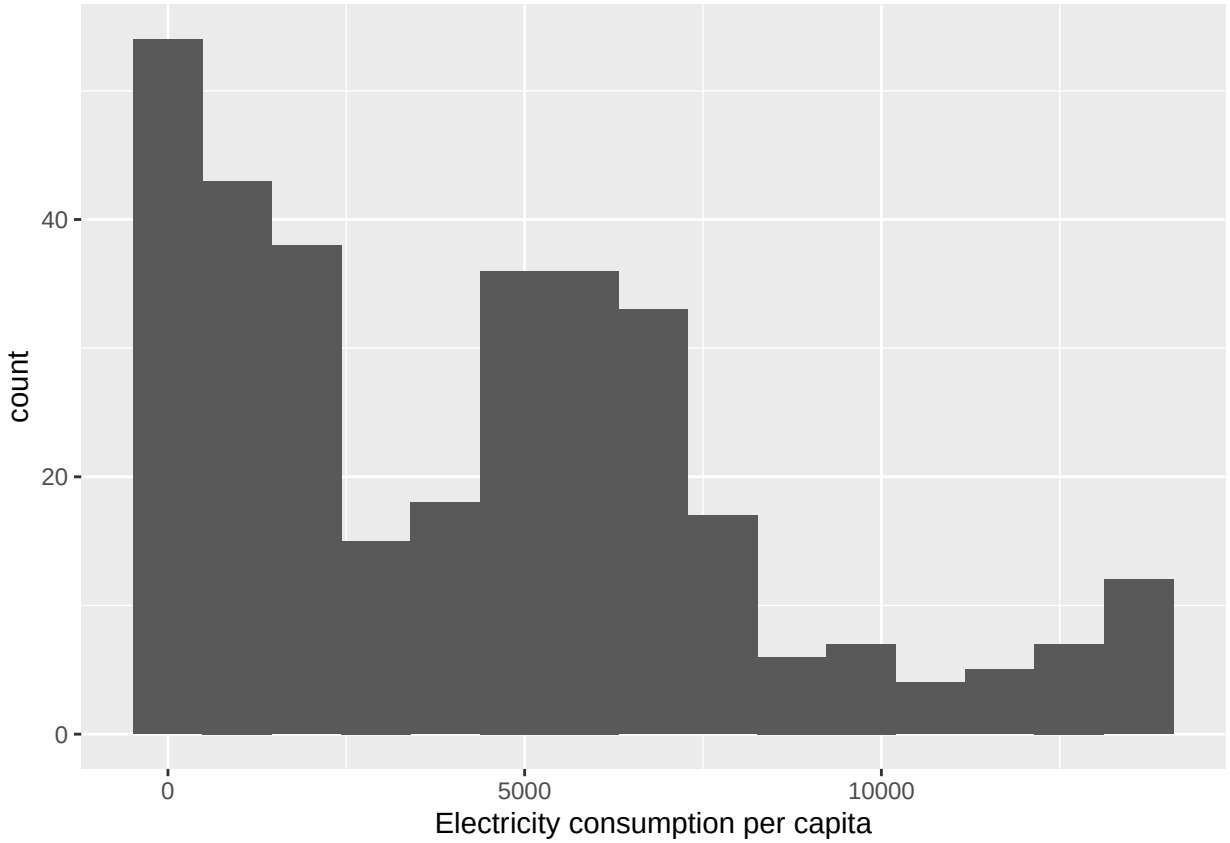

```
p3
```

```
## Warning: Removed 1181 rows containing non-finite values (‘stat_bin()‘).
```




```
p4
```

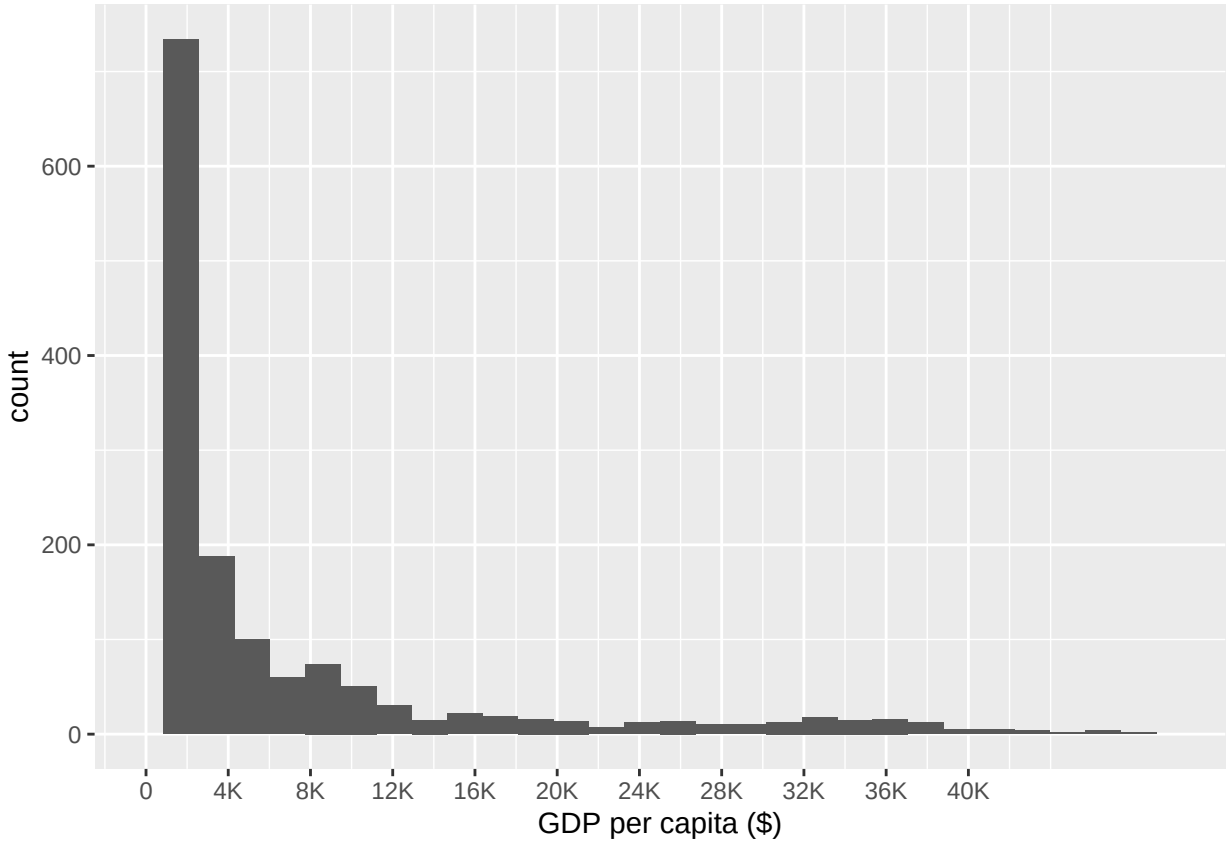
```
## Warning: Removed 1181 rows containing non-finite values (‘stat_bin()‘).
```



Scales

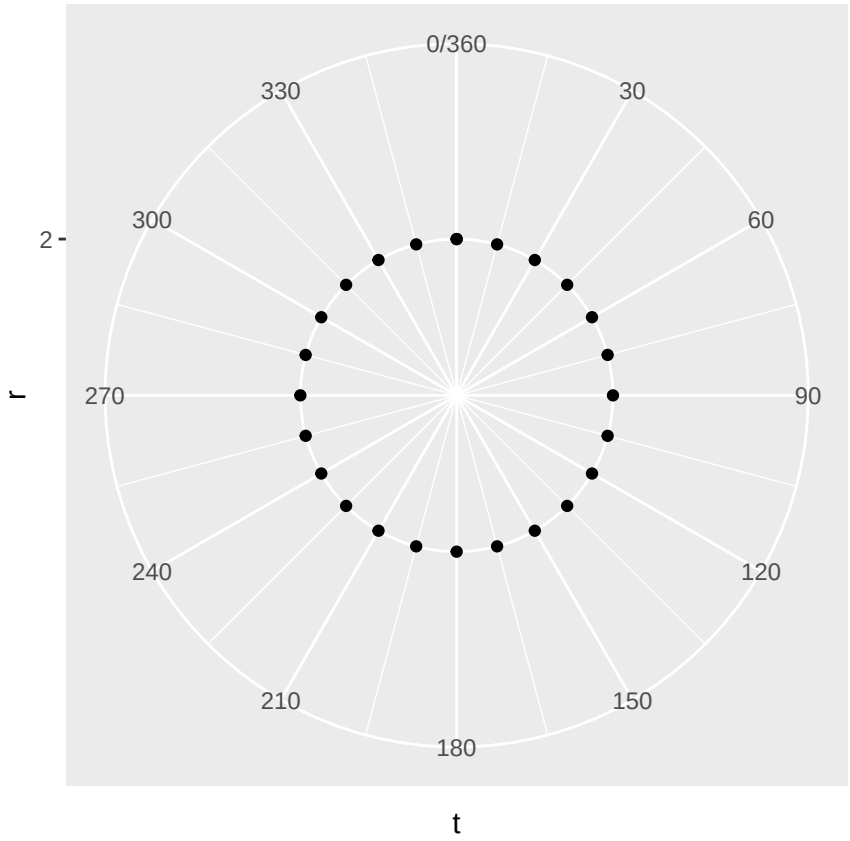
```
p1 <- ggplot(df, aes(x=gdp_per_capita))
p2 <- p1 + geom_histogram()
p3 <- p2 + scale_x_continuous(name='GDP per capita ($)',
                              limits=c(0, 50000),
                              breaks=seq(0, 40000, 4000),
                              labels=c('0', '4K', '8K', '12K', '16K', '20K',
                                       '24K', '28K', '32K', '36K', '40K'))
p3
```

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

## Warning: Removed 7 rows containing non-finite values (`stat_bin()`).

## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



Polar coordinates

```
t <- seq(0, 360, by=15)
r <- 2
qplot(r, t) +
coord_polar(theta="y") +
scale_y_continuous(breaks=seq(0, 360, 30))
```


0/360
30
60
90
120
150
180
210
240
270
300
330
2
r
t