Graphics with ggplot2 (solution)

Data Science Lab, University of Copenhagen

01 September, 2021

Loading the core tidyverse packages and the 'readxl' package for data import from .xlsx.

```
library(tidyverse)
library(readxl)
```

Importing the climate data from **climate.xlsx**[1]. (Change the path to the Excel file below so that it matches the path to the file saved on your own computer, or use *Import Dataset* in RStudio to obtain the relevant code.)

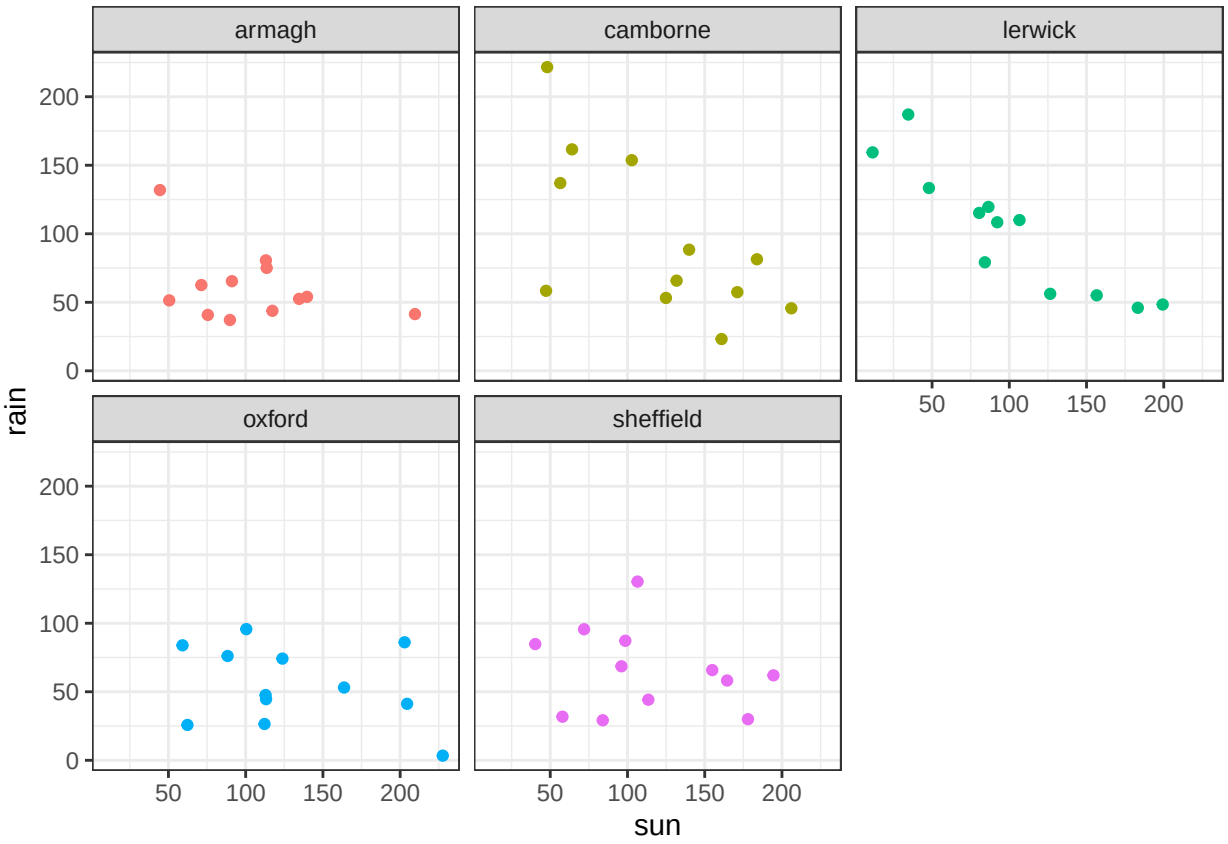
```
climate <- read_excel("climate.xlsx")
climate
```

```
## # A tibble: 60 x 7
##    station  year month    af  rain   sun device
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
##  1 armagh   2016     1     5 132.   44.5 Campbell Stokes
##  2 armagh   2016     2    10  62.6  71.3 Campbell Stokes
##  3 armagh   2016     3     4  43.8 117.  Campbell Stokes
##  4 armagh   2016     4     5  54   140.  Campbell Stokes
##  5 armagh   2016     5     0  41.4 210.  Campbell Stokes
##  6 armagh   2016     6     0  75.1 114.  Campbell Stokes
##  7 armagh   2016     7     0  80.6 113.  Campbell Stokes
##  8 armagh   2016     8     0  52.5 135.  Campbell Stokes
##  9 armagh   2016     9     0  65.4  91.1 Campbell Stokes
## 10 armagh   2016    10     0  37.1  89.8 Campbell Stokes
## # ... with 50 more rows
```

Scatter plot I

```
ggplot(climate, aes(x = sun, y = rain, colour = station)) +
  geom_point() +
  facet_wrap(~station) +
  theme_bw() +
  theme(legend.position = "none")
```

[1] Contains public sector information licensed under the Open Government Licence v3.0.



Graphic files

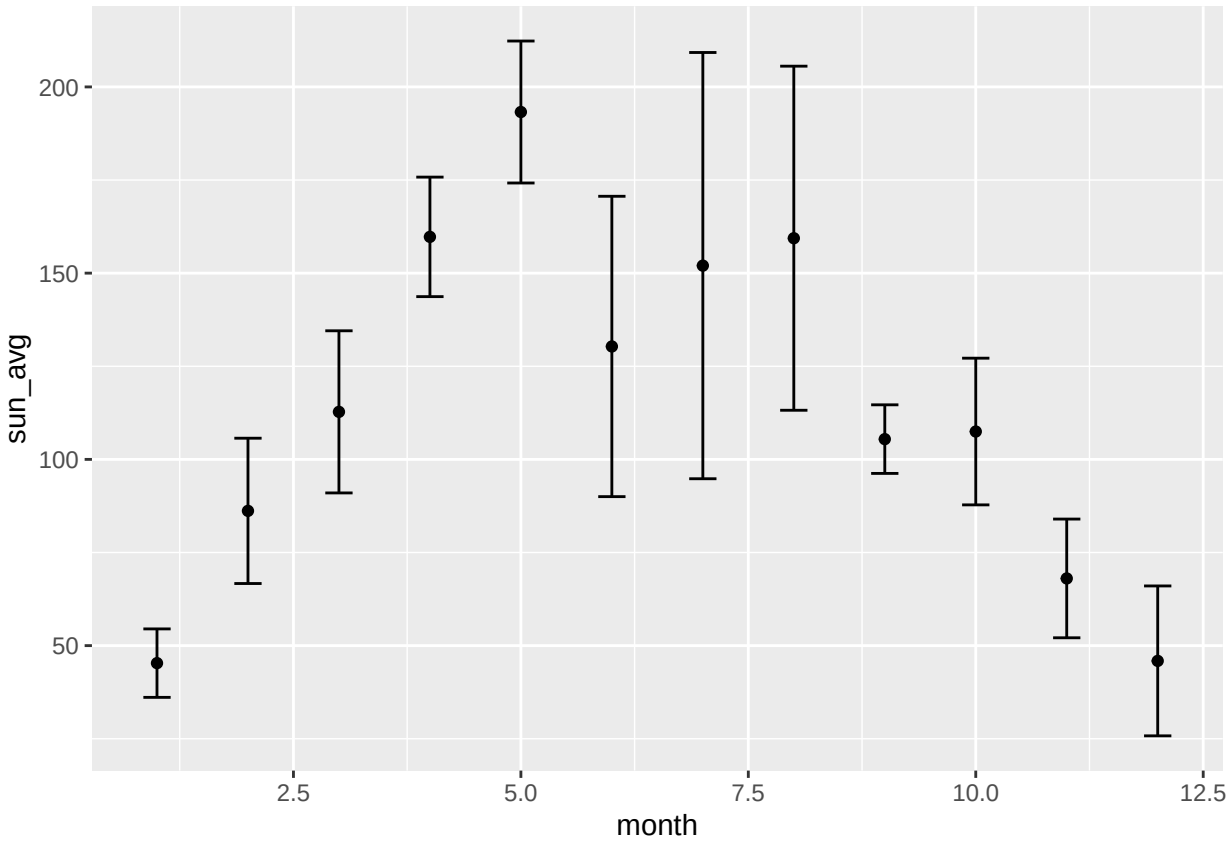
```
ggsave(file="weather.jpeg")

ggsave(file="weather.png",width=10,height=8,units="cm")
```

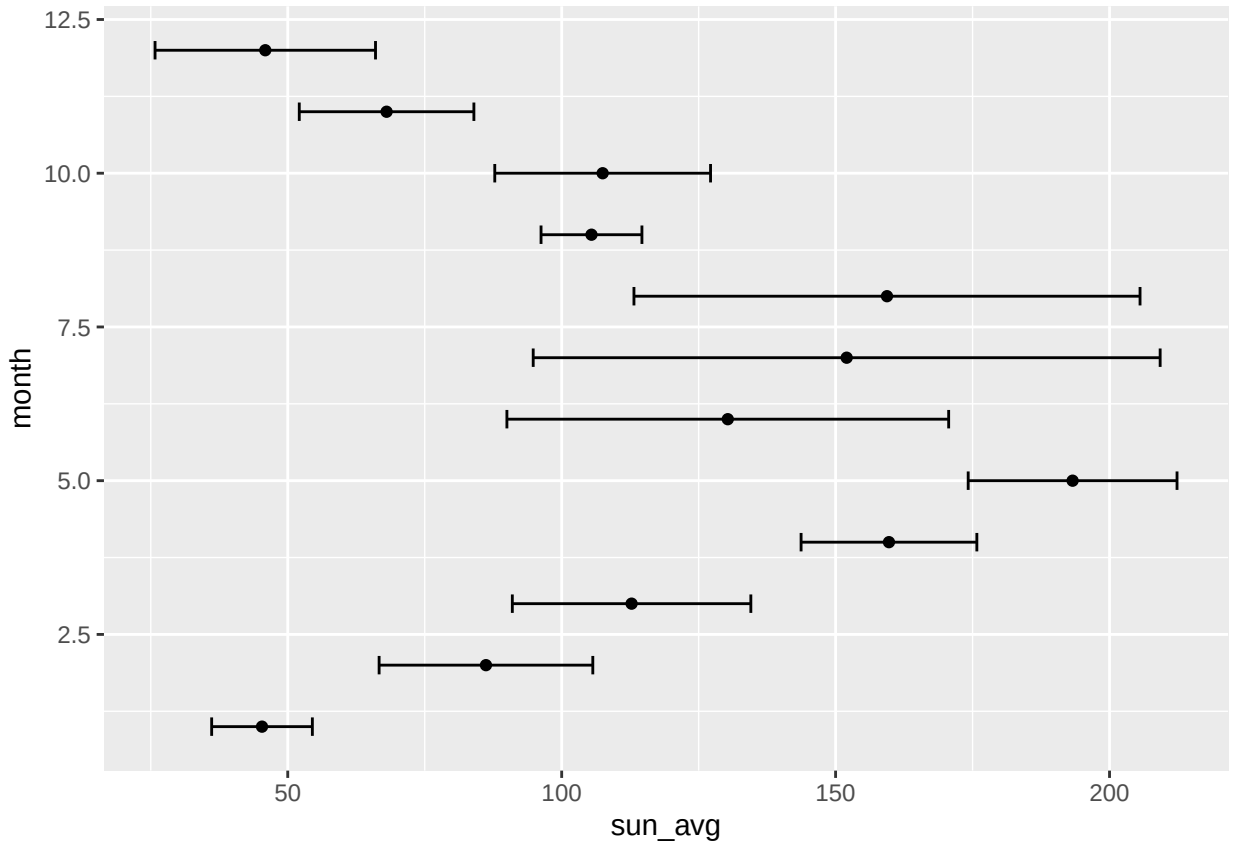
Scatter plot II

```
summary_stats <- climate %>%
  group_by(month) %>%
  summarize(sun_avg = mean(sun), sun_sd = sd(sun))

p <- ggplot(summary_stats, aes(x = month, y = sun_avg)) +
  geom_point() +
  geom_errorbar(aes(ymin = sun_avg - sun_sd, ymax = sun_avg + sun_sd), width = 0.3)
p
```




```
p + coord_flip()
```

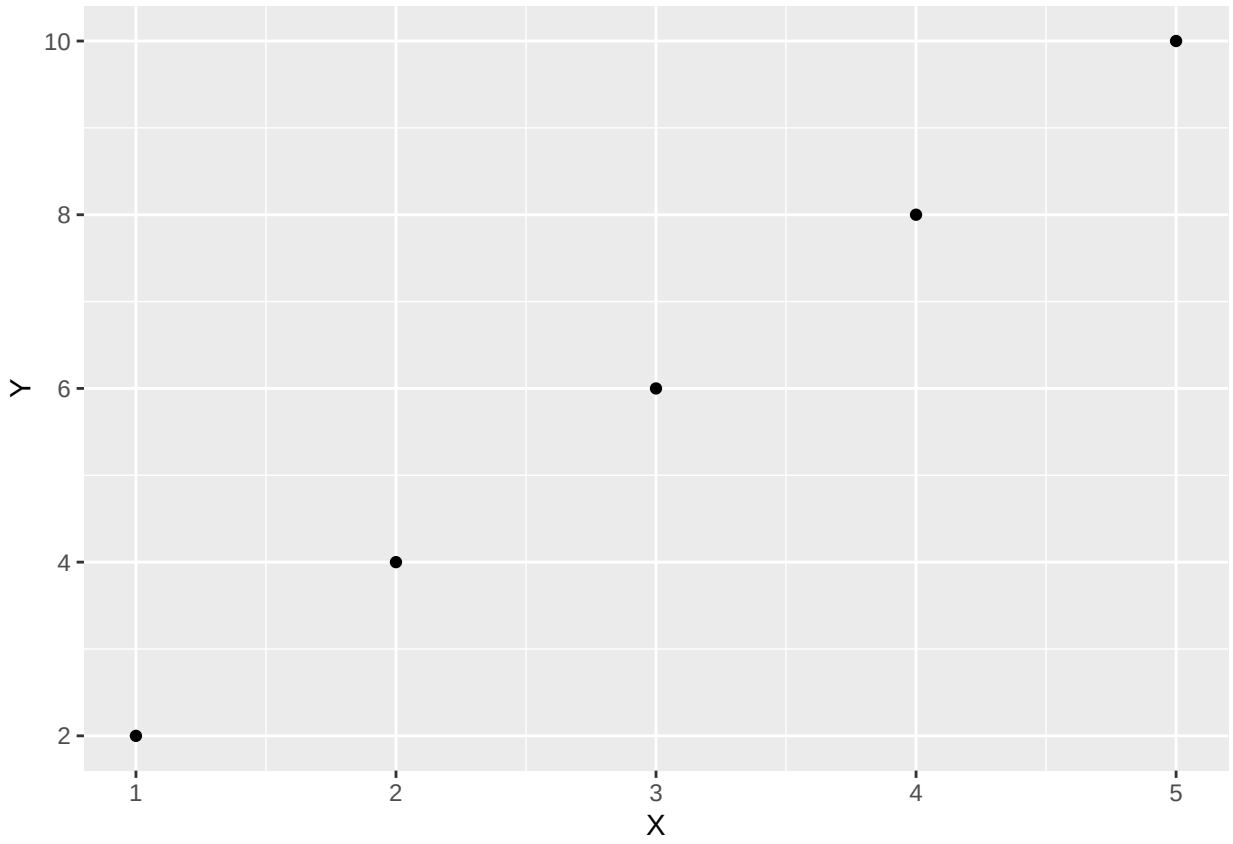


Scatter plot III

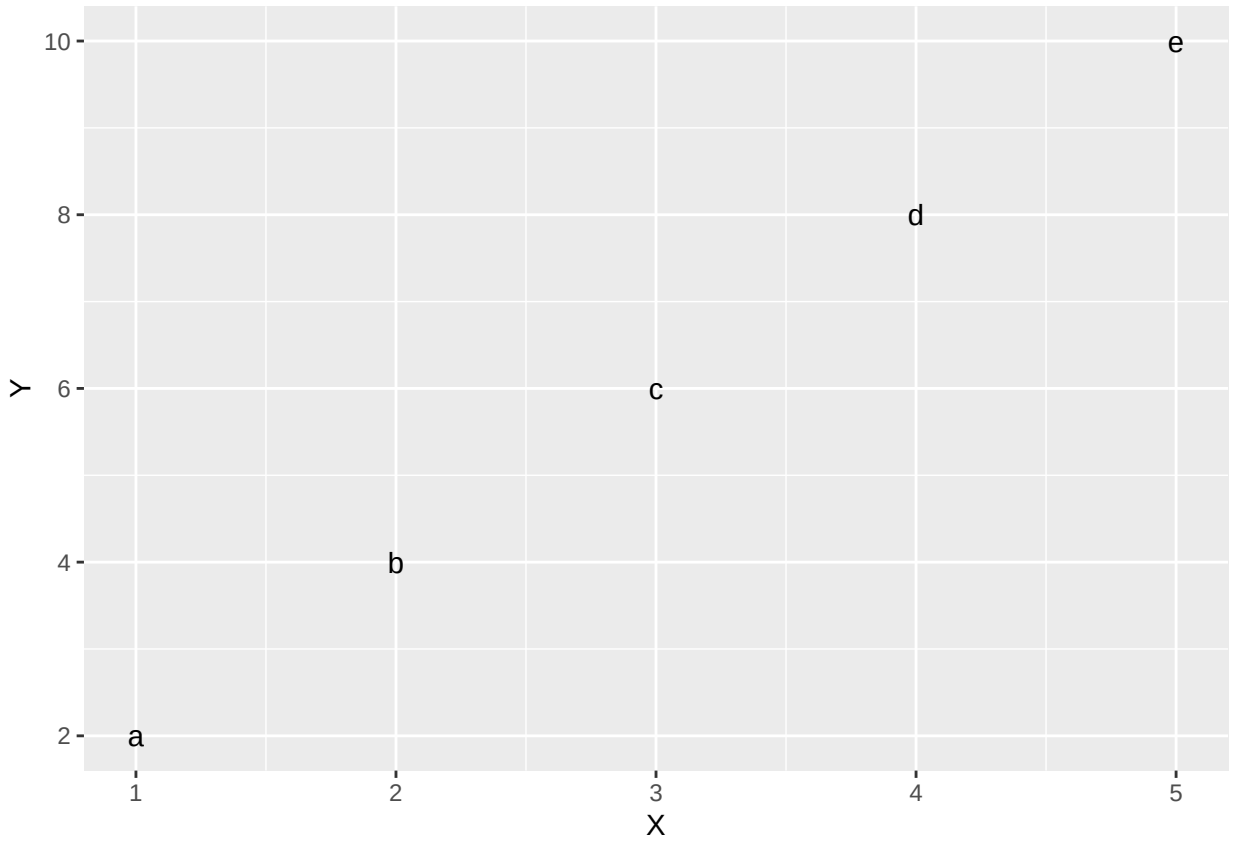
```
example_data <- tibble(X = 1:5, Y = 2*X, Z = letters[1:5])
example_data
```

```
## # A tibble: 5 x 3
##       X     Y Z
##   <int> <dbl> <chr>
## 1     1     2 a
## 2     2     4 b
## 3     3     6 c
## 4     4     8 d
## 5     5    10 e
```

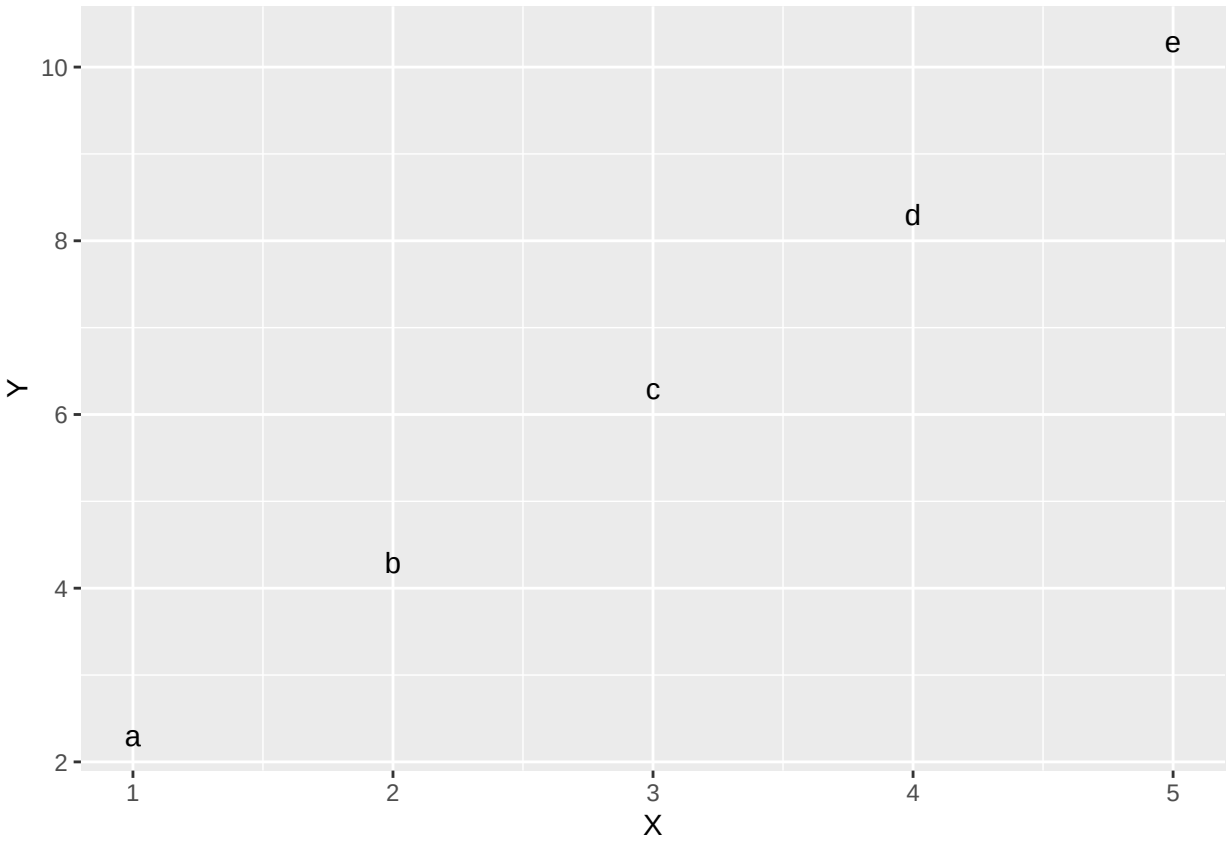
```
ggplot() +
  geom_point(mapping = aes(x = X, y = Y), data = example_data)
```




```
ggplot() +
  geom_text(mapping = aes(x = X, y = Y, label = Z), data = example_data)
```




```
ggplot() +
  geom_text(mapping = aes(x = X, y = Y, label = Z), data = example_data, nudge_y = 0.3)
```

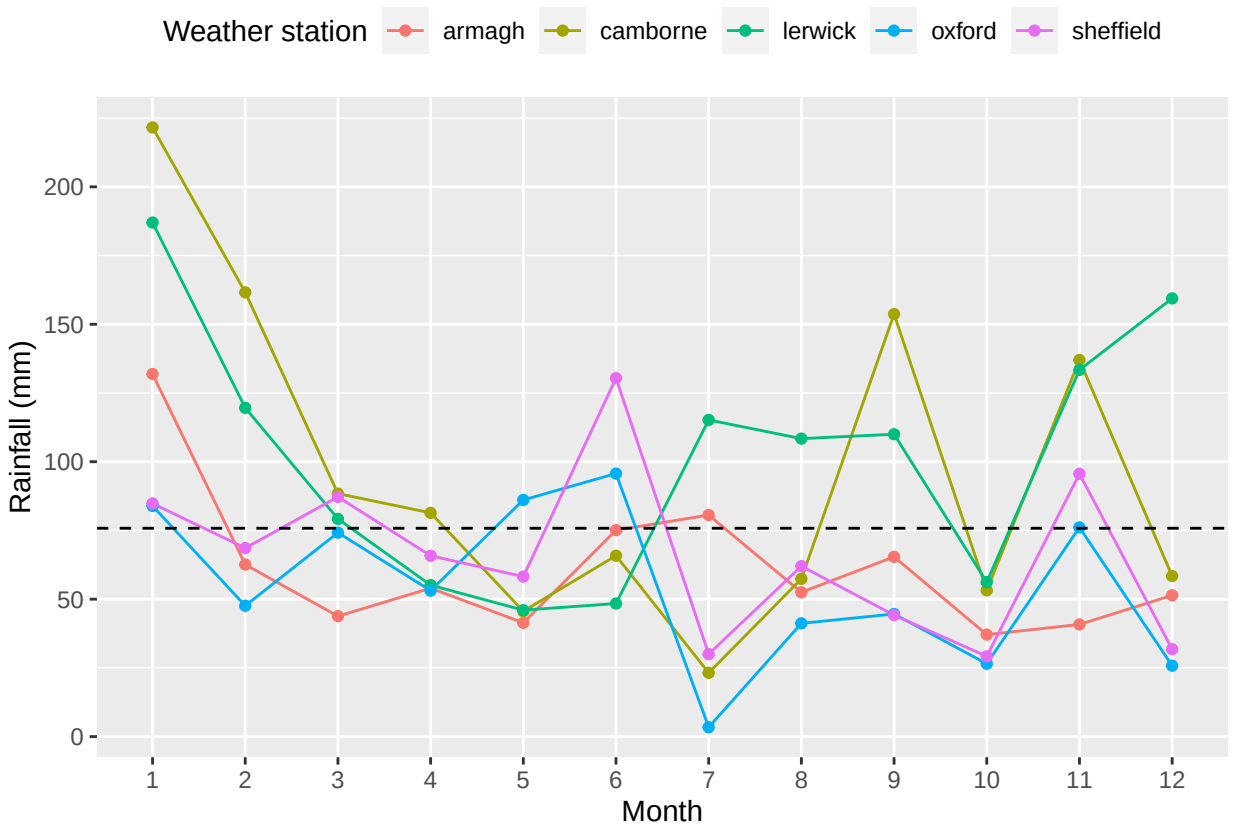


Line plot

The line plot was not discussed at the presentation, which makes this question slightly more difficult. So you need to figure out by yourself that `geom_line` is used to make lines. Furthermore, you need the `group` aesthetic to group together the data points to be connected by lines. Thus:

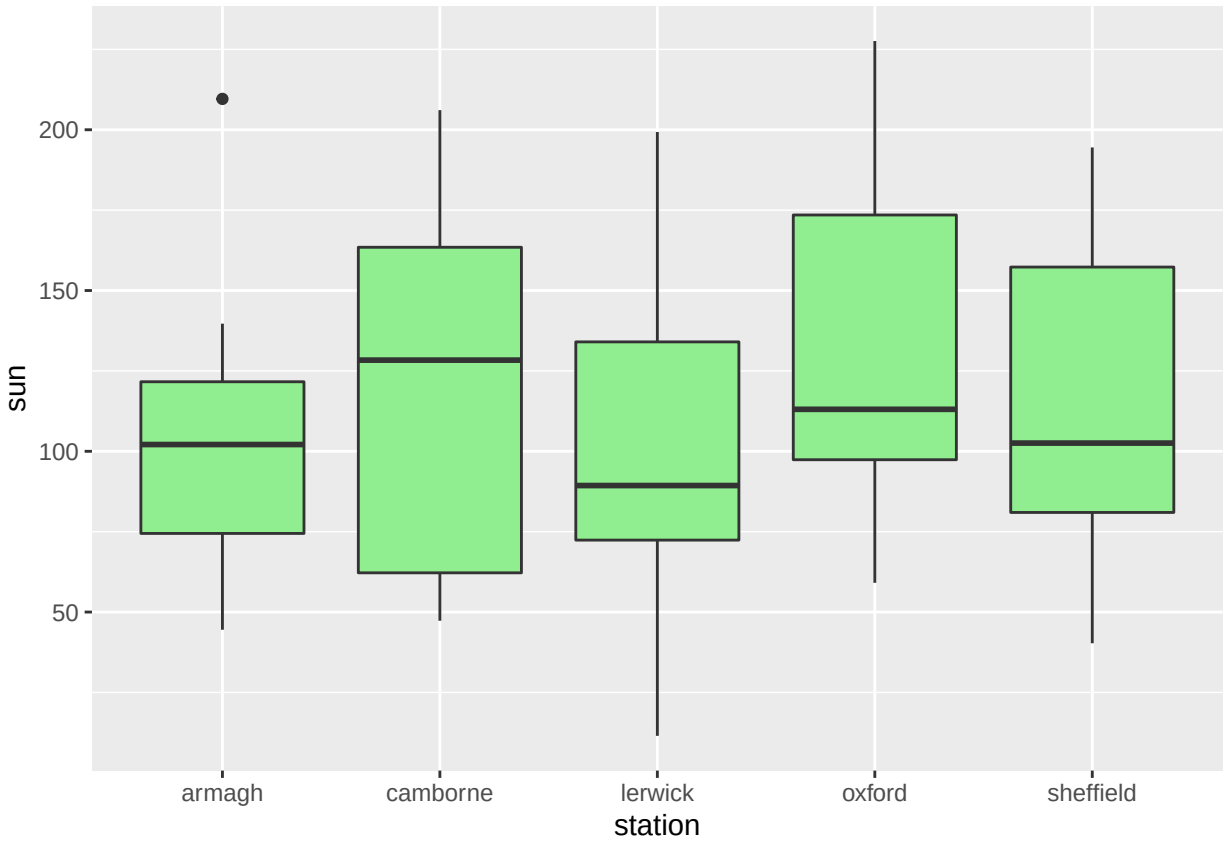
```
climate <- mutate(climate, month = factor(month))

ggplot(climate, aes(x = month, y = rain, group = station, colour = station)) +
  geom_line() +
  geom_point() +
  theme(legend.position = "top") +
  geom_hline(yintercept = mean(climate$rain), linetype = "dashed") +
  labs(x = "Month", y = "Rainfall (mm)", colour = "Weather station")
```

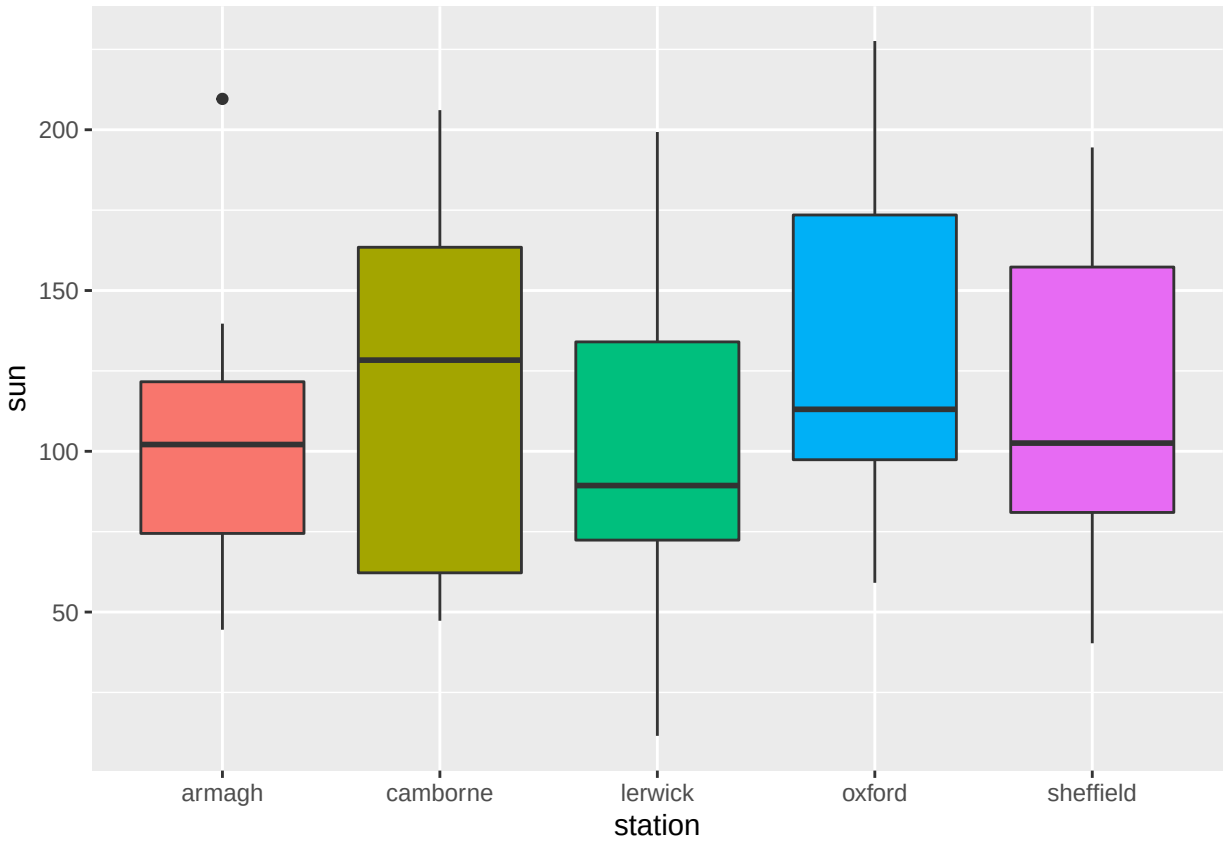


Box plot I

```
ggplot(climate, aes(x = station, y = sun)) +
  geom_boxplot(fill = "lightgreen")
```

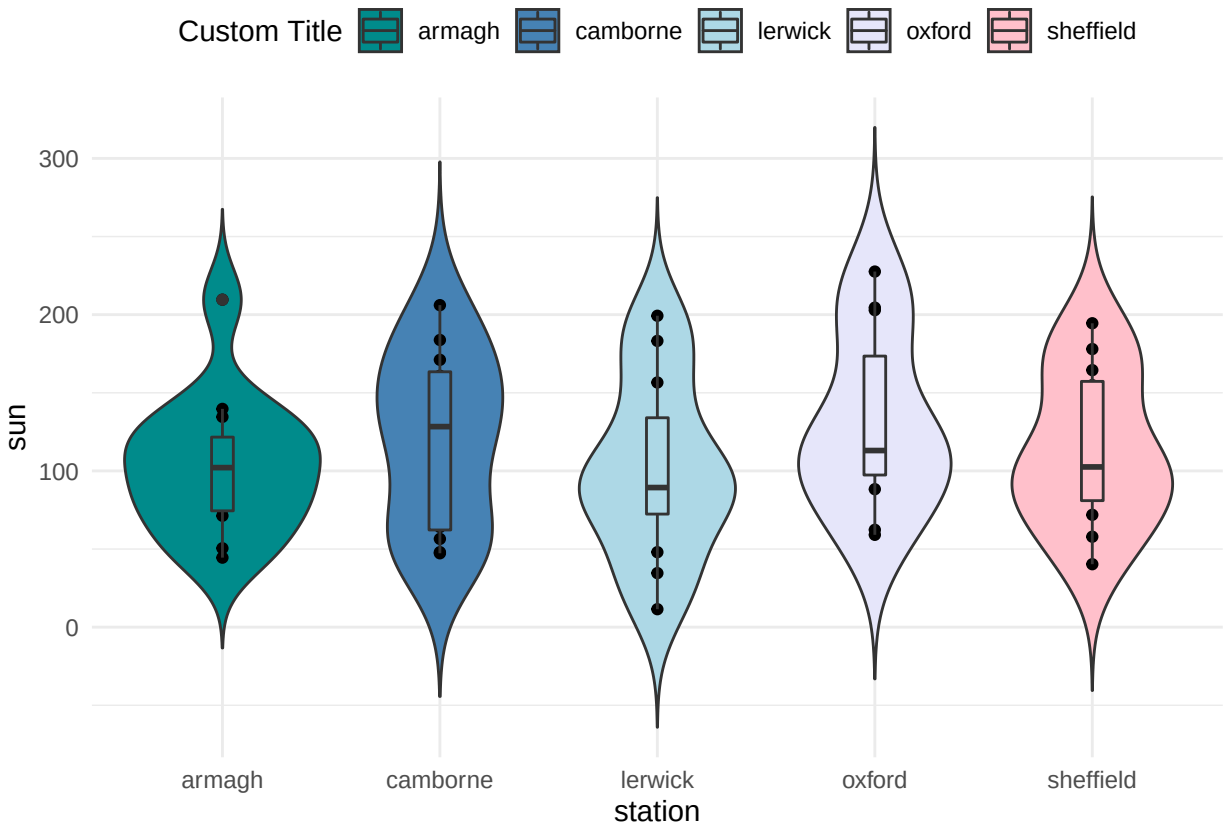



```
ggplot(climate, aes(x = station, y = sun)) +
  geom_boxplot(aes(fill = station)) +
  theme(legend.position = "none")
```



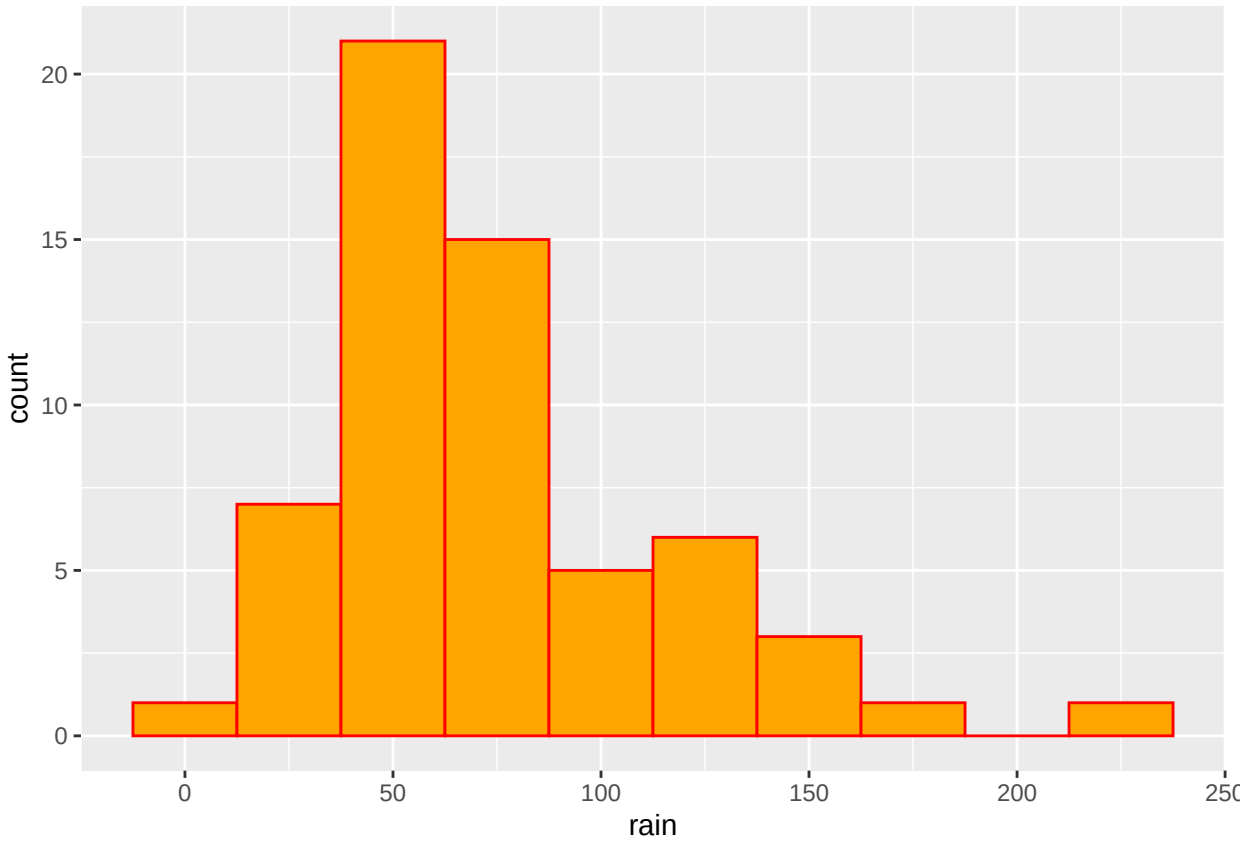
Box plot II - Aesthetics

```
ggplot(climate, aes(x=station, y=sun, fill=station)) + geom_violin(trim=FALSE) +
  geom_point() + theme_minimal() + geom_boxplot(width=.1) +
  scale_fill_manual(values = c("darkcyan","steelblue", "lightblue", "lavender", "pink")) +
  labs(fill = "Custom Title") + theme(legend.position="top")
```



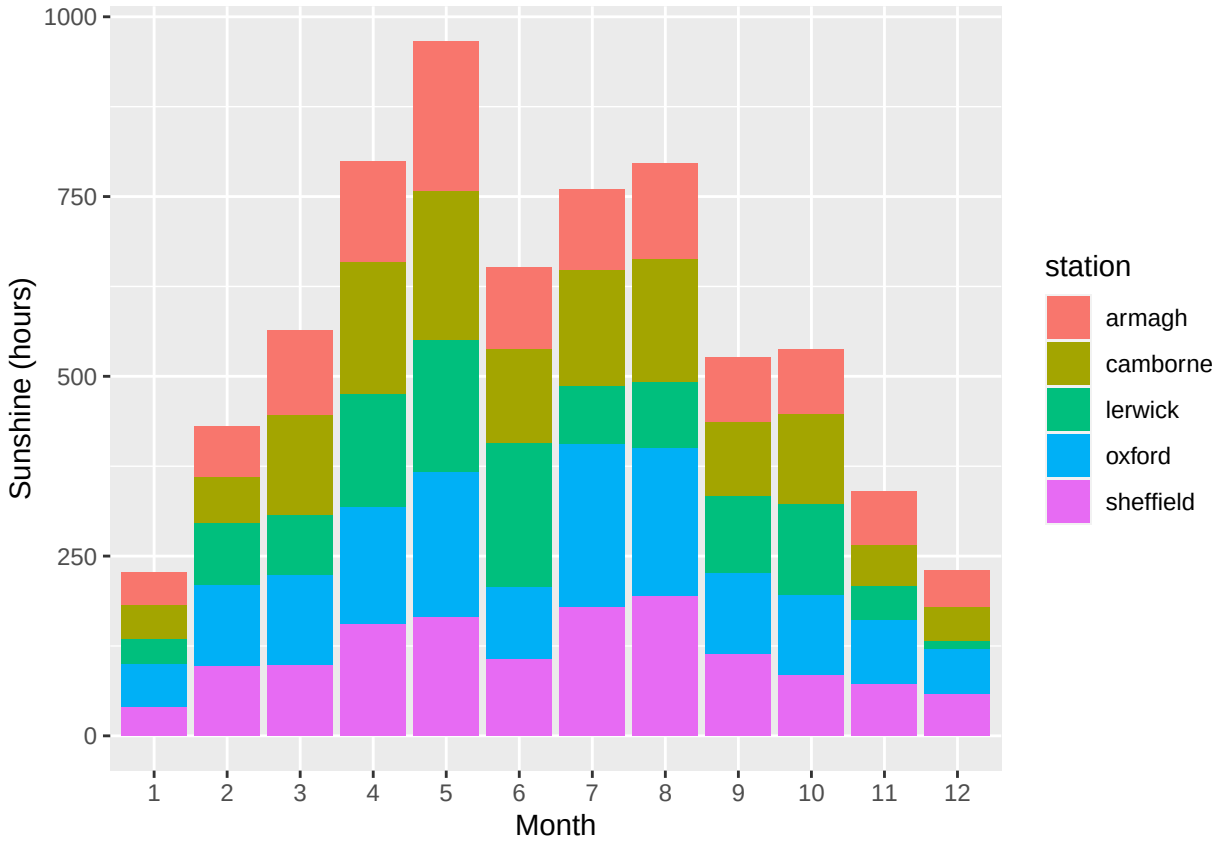
Histogram

```
ggplot(climate, aes(x = rain)) +
  geom_histogram(binwidth = 25, colour = "red", fill = "orange")
```



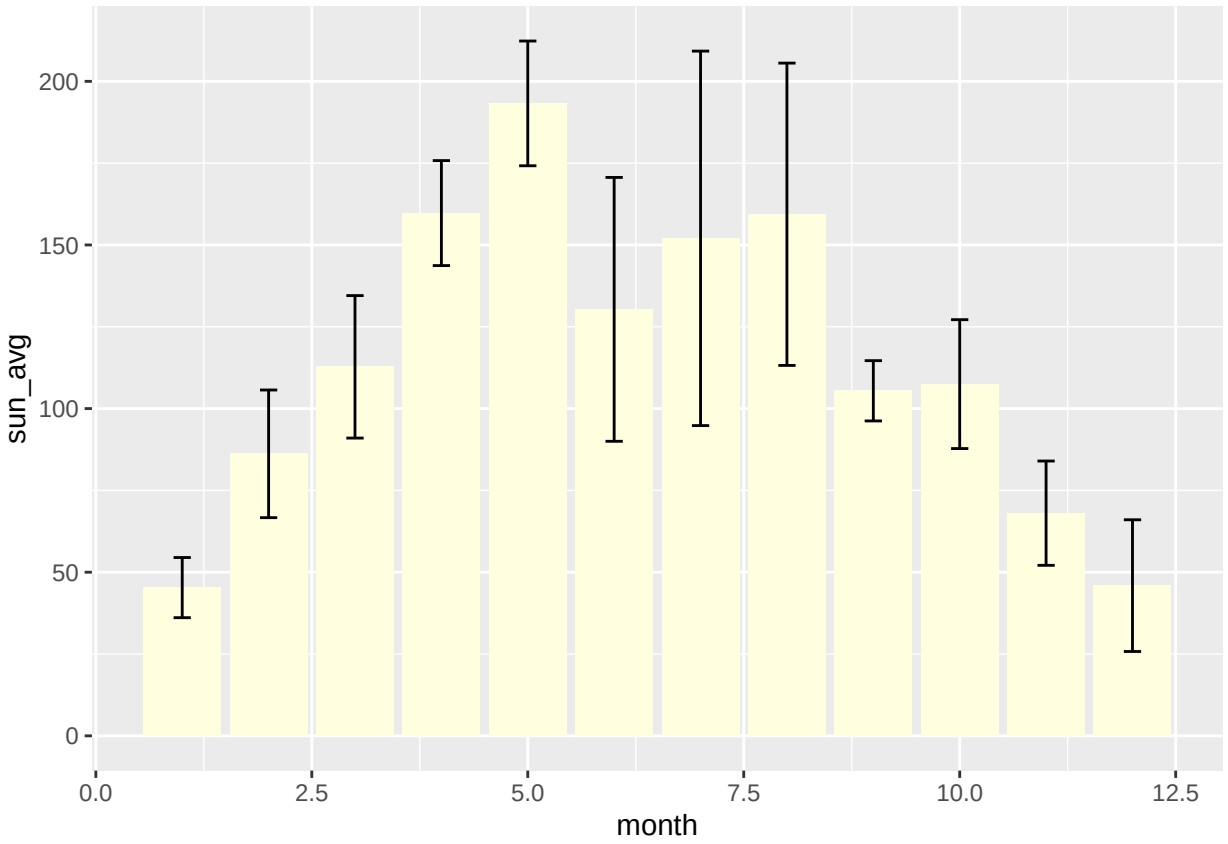
Bar chart I

```
ggplot(climate, aes(x = month, y = sun, fill = station)) +
  geom_col() +
  labs(x = "Month", y = "Sunshine (hours)", colour = "Weather station")
```



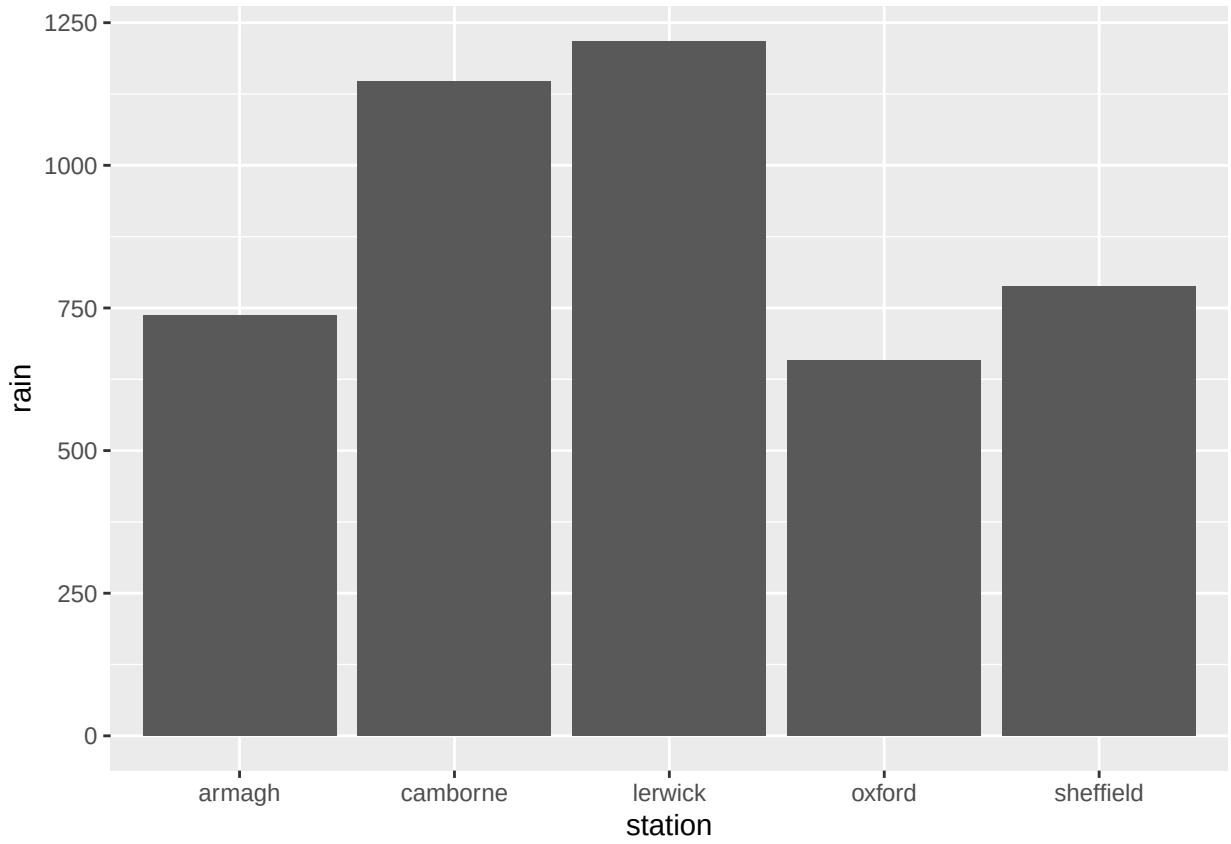
Bar chart II

```
ggplot(summary_stats, aes(x = month, y = sun_avg)) +
  geom_col(fill = "lightyellow") +
  geom_errorbar(aes(ymin = sun_avg - sun_sd, ymax = sun_avg + sun_sd), width = 0.2)
```



Bar chart III

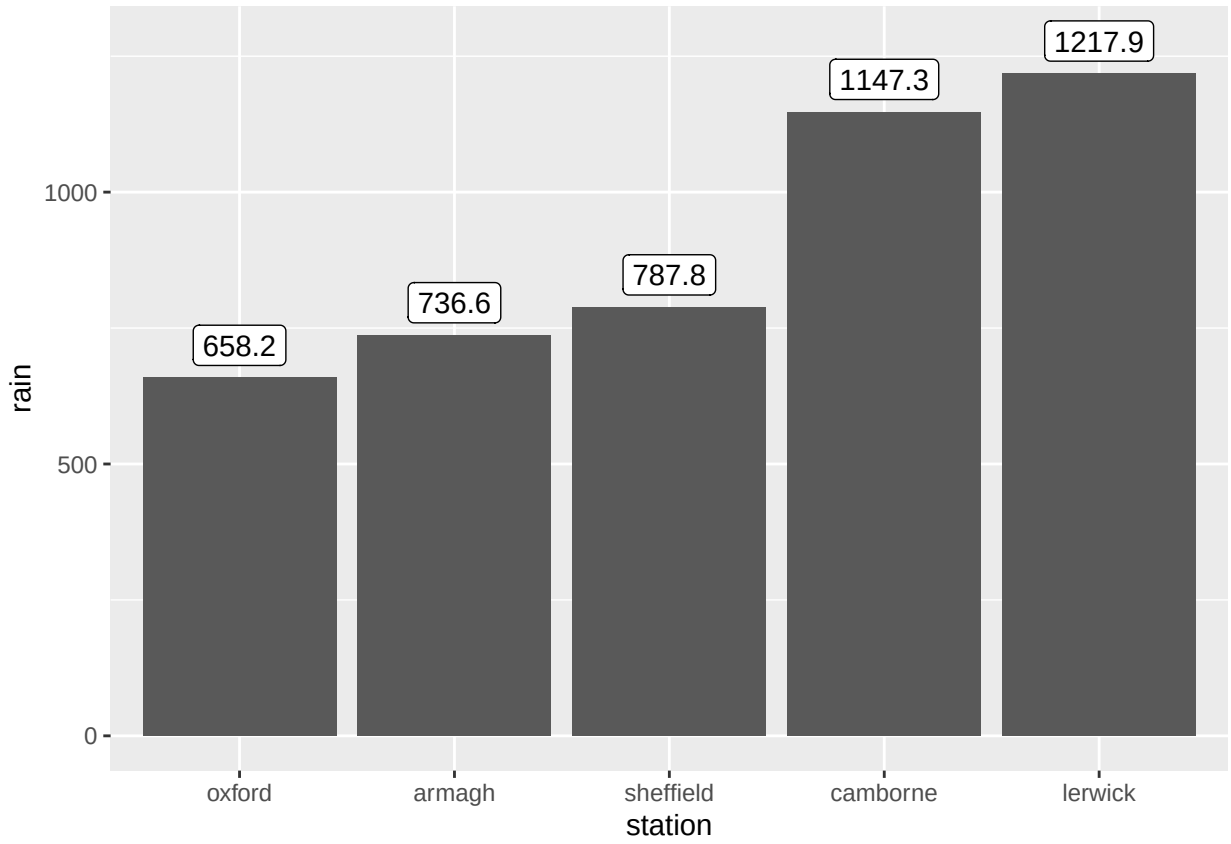
```
ggplot(climate, aes(x = station, y = rain)) +
  geom_col()
```




```
annual_rain <-
  climate %>%
  group_by(station) %>%
  summarize(rain = sum(rain)) %>%
  arrange(rain)

climate <- mutate(climate, station = factor(station, levels = annual_rain$station))

ggplot(climate, aes(x = station, y = rain)) +
  geom_col() +
  geom_label(mapping = aes(x = station, y = rain, label = rain), data = annual_rain, nudge_y = 60)
```


1217.9
1147.3
787.8
736.6
658.2
1000
500
0
rain
oxford
armagh
sheffield
camborne
lerwick
station