Bar charts are best used to show 'counts' rather than 'proportions'.

One of the most common types of charts, bar charts display the quantities of qualitative or categorical data.

Bar charts are **most** suitable when:

- **you have categorical, comparative data**
- **your data is classified into nominal or ordinal categories.**

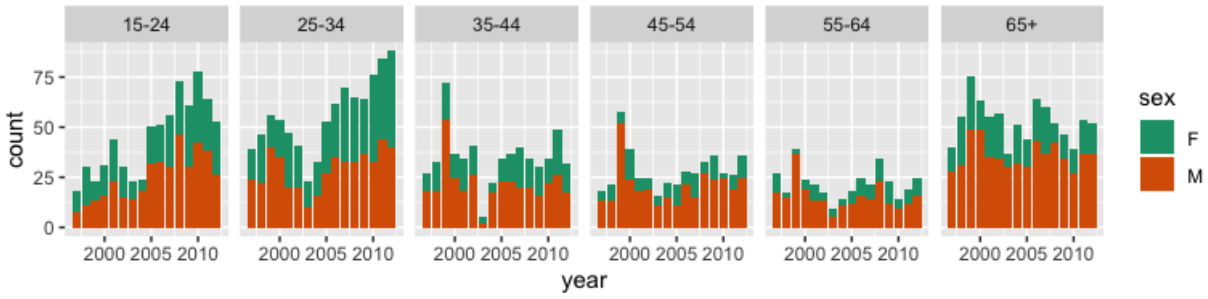
Bar charts are **least** suitable when:

- **you have many categories**
- **you have more data series being added.**

To find out more about choosing a suitable chart to visualise your data, you might like to read How to select among different chart types.

Incidents of tuberculosis

Consider the following sample chart, which shows incidents of tuberculosis (TB).

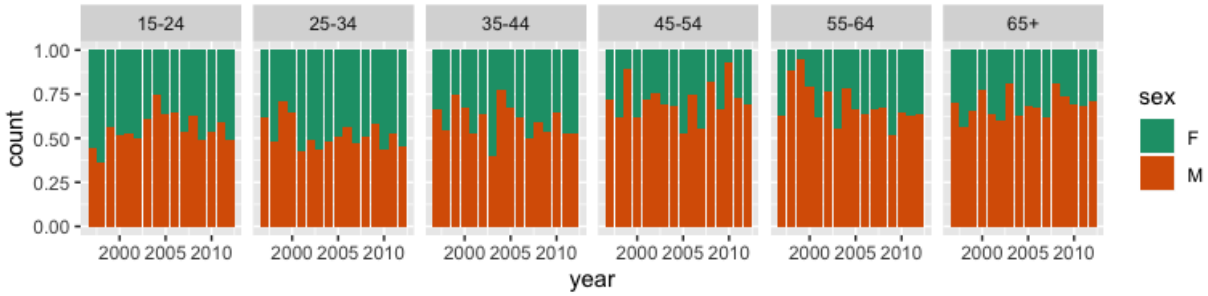


What do we learn from this chart?

The focus of the chart is on **counts** in each category. It shows that counts are different across **ages** and **years**: counts tend to be lower in middle age (45-64).

It also shows that in the year 1999, there was a bit of a TB outbreak in most age groups, with numbers doubling or tripling compared to other years. The TB Incidence has been increasing among younger age groups in recent years.

How is the sample chart on this step different from the 100% chart you produced?



Give it a go!

Continue to develop your skills in developing graphic plots by creating your own bar chart.

If you haven't already, open RStudio on your computer and load the TB data. For this exercise you will also need to load the **tidyverse** package in RStudio on your computer:

```
library(tidyverse)
tb <- read_rds("data/tb_tidy.rds")
tb_au <- filter(tb,
                country == "Australia",
                !is.na(age_group))
```

Once you've installed the data and **tidyverse**, return to this step and then follow along.

Examine your bar chart. Are the variable mappings the same or different compared to the 100% chart you produced on the previous step, and have the colours changed?

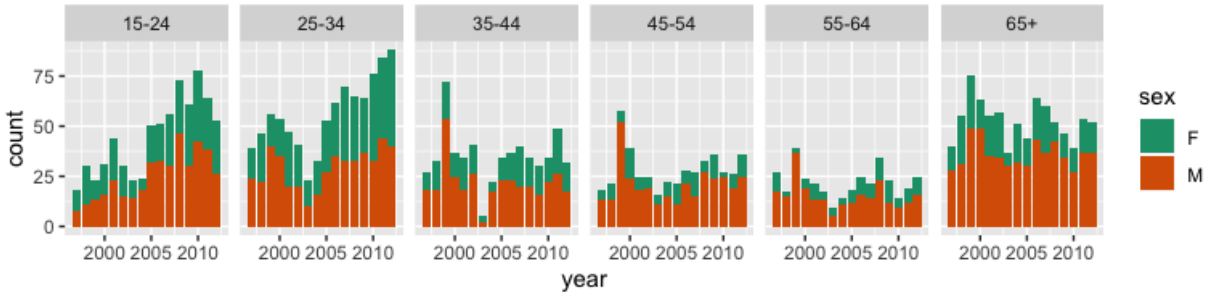
In the bar chart we have the same facets, colours and aesthetics, but what has changed in the way the bars have been positioned?

Instead of filling them up to 100%, they're **stacked** based on their counts.

Copy and run

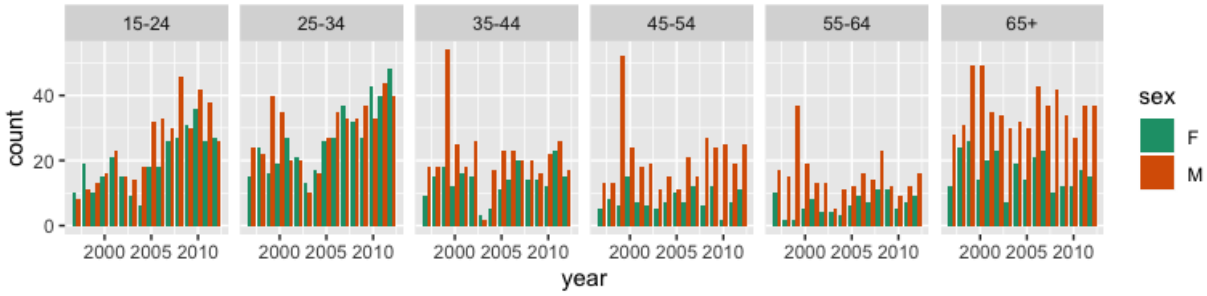
For your own bar chart, you can achieve this using **ggplot2** by changing the position argument in **geom_bar()** to **stack**, with the following code chunk:

```
p <- ggplot(tb_au, aes(x = year, y = count, fill = sex)) +
  geom_bar(stat = "identity", position = "stack") +
  facet_grid(~ age_group) +
  scale_fill_brewer(palette="Dark2")
p
```



Going from 'stacked' to 'side-by-side'

What if you wanted to show the counts in each age category but with respect to **sex**, as shown in the following chart?



What can you learn from this chart?

The focus is now on counts by sex, and shows predominantly male incidence of TB. You can also see that incidence among males relative to females is from middle age onwards.

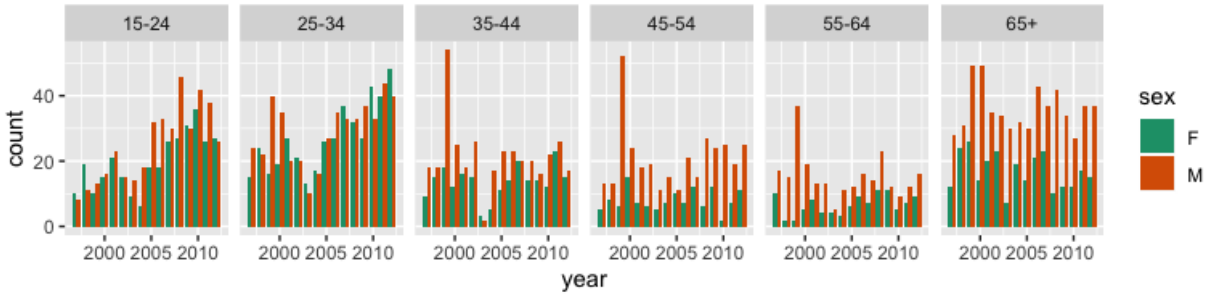
There appears to be similar incidence between males and females in younger age groups.

Again, let's think about what's changed from your previous charts. The colours, aesthetics and facets have remained the same, it's just the **position** of the bars that have been altered.

Copy and run

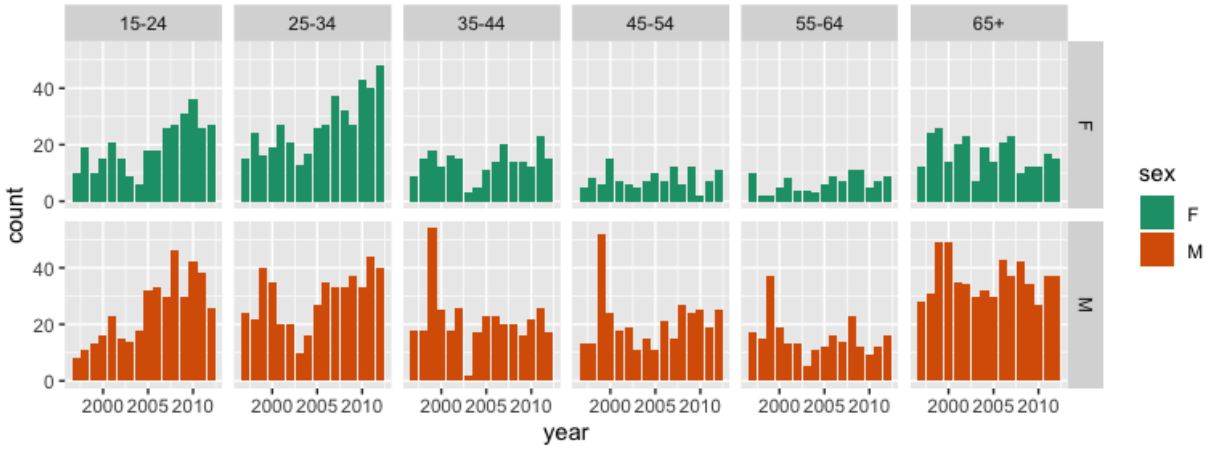
For your own bar chart, you can achieve this using **`ggplot2`** by changing the position argument in **`geom_bar()`** to **`dodge`**, with the following code chunk:

```
p <- ggplot(tb_au, aes(x = year, y = count, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(~ age_group) +
  scale_fill_brewer(palette="Dark2")
p
```



Going from 'side-by-side' to 'separate bars'

You also might be interested in looking at how the incidence changes within a sex over time and age groups, as shown in the following chart:



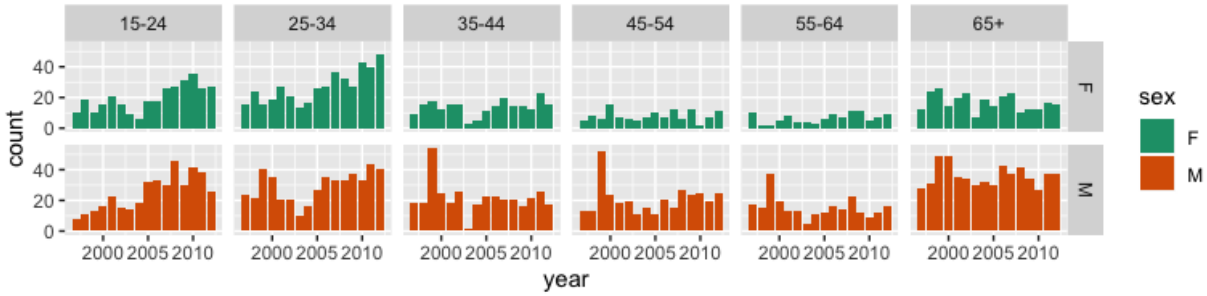
What do you learn?

It's now easier to focus separately on males and females. You can see that the 1999 outbreak of TB mostly affected males. Incidences of TB cases continue to grow for females in the 25-34 age group while stabilising for males in the same age group.

What changes in the code?

You may have noticed you are facetting on an additional variable: **sex**. You also no longer require a position argument, since there is only one bar per category.

```
p <- ggplot(tb_au, aes(x = year, y = count, fill = sex)) +
  geom_bar(stat = "identity") +
  facet_grid(sex ~ age_group) +
  scale_fill_brewer(palette="Dark2")
p
```



Tell us how you went

Within the **Comments** share with other learners your experience of creating your own bar chart using the grammar, and the following:

- **How does the structure of the chart allow you to explore the TB data?**

Next, you will learn how pie charts are really bar charts, but just expressed in a different coordinate system. When you're ready, make your way to the next step.