forcats practical

Jumping Rivers

First we'll load the data and the relevant packages

```
library(forcats)
library(tidyverse)
data(okcupid, package = "jrTidyverse")
```

This is the okcupid data from yesterday's course. We're going to specifically focus on how drinking affects income.

Question 1

a) The column `drinks` corresponds to a persons answer about their drinking habits. What is the average income of each group of drinking habits? Save this as a data frame called `drinks_in` and have the column containing the average income called `av_in` Hint: use `group_by() then summarise()`

b) We can plot the average incomes using **ggplot2**

```
drinks_in %>%
  ggplot(aes(x = drinks, y = av_in)) +
  geom_point()
```

c) Previously we saw how to rename factors using `fct_recode()`. However, this will not work with missing values i.e. `NA`'s. A function that will is `fct_explicit_na()`. Try running

```
x = c(1,2,3,NA)
(y = factor(x))
```

Notice how the NA isn't included in the factors?

```
fct_explicit_na(y, "unknown")
```

That will rename the NA factors as "Unknown". Before plotting, use `mutate()` and `fct_explicit_na()` to rename the missing values to something more appropriate.

d) Before plotting, reorder the points from lowest average income to highest.
Hint: use `mutate()` and `fct_reorder()`

e) Before plotting, instead of ordering the points from lowest income to highest, order them from people who drink least to people who drink most. Put "Unknown"" where you deem appropriate.
Hint: use `fct_relevel()`

f) Go back to before we summarised the average income of each group. Summarise the groups average income in the same way, but this time collapse "not at all" and "rarely" into "low", "socially" and "often" into "medium" and then "very often" and "desperately" into "high".