Titanic Exercise: Code

```
>library(ggplot2)

> # titanic is available in your workspace

> # 1 - Check the structure of titanic
> str(titanic_clean)

Classes ‘spec_tbl_df’, ‘tbl_df’, ‘tbl’ and 'data.frame':	1309 obs. of  15 variables:
 $ pclass        : num  1 1 1 1 1 1 1 1 1 1 ...
 $ survived      : num  1 1 0 0 0 1 1 0 1 0 ...
 $ name          : chr  "Allen, Miss. Elisabeth Walton" "Allison, Master. Hudson Trevor" "Allison,
Miss. Helen Loraine" "Allison, Mr. Hudson Joshua Creighton" ...
 $ sex           : chr  "female" "male" "female" "male" ...
 $ age           : num  29 0.917 2 30 25 ...
 $ sibsp         : num  0 1 1 1 1 0 1 0 2 0 ...
 $ parch         : num  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket        : chr  "24160" "113781" "113781" "113781" ...
 $ fare          : num  211 152 152 152 152 ...
 $ cabin         : chr  "B5" "C22 C26" "C22 C26" "C22 C26" ...
 $ embarked      : chr  "S" "S" "S" "S" ...
 $ boat          : chr  "2" "11" "NONE" "NONE" ...
 $ body          : num  NA NA NA 135 NA NA NA NA NA 22 ...
 $ home.dest     : chr  "St Louis, MO" "Montreal, PQ / Chesterville, ON" "Montreal, PQ /
Chesterville, ON" "Montreal, PQ / Chesterville, ON" ...
 $ has_cabin_number: num  1 1 1 1 1 1 1 1 1 0 ...
 - attr(*, "spec")=List of 3
  ..$ cols   :List of 15
  .. ..$ pclass        : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ survived      : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ name          : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ sex           : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ age           : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ sibsp         : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ parch         : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ ticket        : list()
```

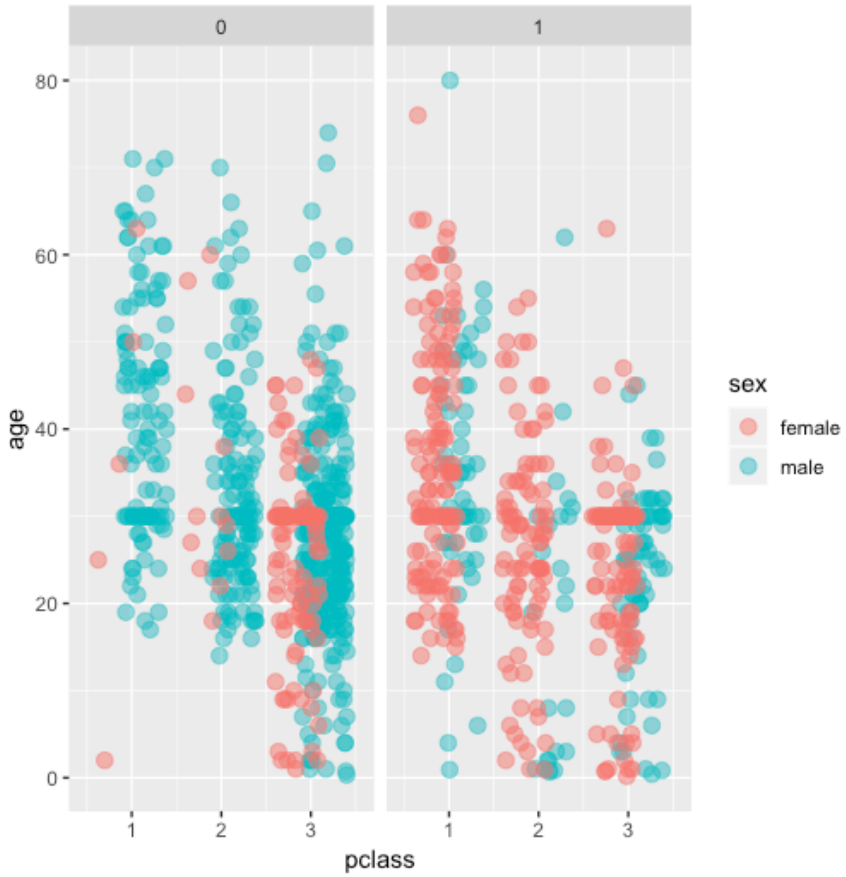
```
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ fare          : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ cabin         : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ embarked      : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ boat          : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ body          : list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  .. ..$ home.dest     : list()
  .. .. ..- attr(*, "class")= chr  "collector_character" "collector"
  .. ..$ has_cabin_number: list()
  .. .. ..- attr(*, "class")= chr  "collector_double" "collector"
  ..$ default: list()
  .. ..- attr(*, "class")= chr  "collector_guess" "collector"
  ..$ skip   : num 1
  ..- attr(*, "class")= chr "col_spec"

> # 2 - Use ggplot() for the first instruction
> ggplot(titanic_clean, aes(x = pclass, fill = sex)) +
+     geom_bar(position = "dodge")

> # 3 - Plot 2, add facet_grid() layer
> ggplot(titanic_clean, aes(x = pclass, fill = sex)) +
+     geom_bar(position = "dodge") + facet_grid(. ~ survived)

> # 4 - Define an object for position jitterdodge, to use below
> posn.jd <- position_jitterdodge(0.5, 0, 0.6)

> # 5 - Plot 3, but use the position object from instruction 4
> ggplot(titanic_clean, aes(x = pclass, y = age, col = sex)) +
+     geom_point(position = posn.jd, size = 3, alpha = 0.5) + facet_grid(. ~ survived)
```



Titanic Exercise: Explanation

To control for overplotting, we used position_jitterdodge. This is primarily used for aligning points generated through geom_point( ) with dodged boxplots. The jitter.width (degree of jitter in x direction) is 0.5. The jitter.height (degree of jitter in y direction) is 0, which is the default. The dodge.width (the amount to dodge in the x direction) is 0.6.

In the geom_point, the size is 3, which refers to the diameter of the points. The alpha is 0.5, which refers to the transparency of the points where 0 is transparent and 1 is opaque.

The facet_grid simply splits up our overall data according to the levels in a categorical factor variable. In the Titanic dataset, the data is split along the *survived* variable, where 0 means the passenger did not survive and 1 means the passenger did survive.

According to the plot, it looks as though the majority of female first- and second-class passengers survived. For female third-class passengers, the survival rate looks relatively even. For male passengers, the majority of non-survivors were 40 or younger and third-class. For men who did survive, most were 20- to 40-years-old. It appears as though more third-class male passengers survived than second-class passengers. However, there were 322 more third-class than second-class male passengers in total.