Practical 2: Solutions

Jumping Rivers

First we must load the **tidyverse**

```
library("tidyverse")
```

Question 1 - Cocktails!

We've put together a small list containing the ingredients of some classic cocktails

```
data(cocktails, package = "jrTidyverse2")
```

a) How many cocktails are in the list?

```
str(cocktails)
## List of 6
##  $ pina_colada         : chr [1:3] "rum" "coconut cream" "pineapple"
##  $ bloody_mary         : chr [1:7] "vodka" "tomato juice" "lemon" "tabasco sauce" ...
##  $ long_island_iced_tea: chr [1:7] "vodka" "gin" "tequila" "rum" ...
##  $ espresso_martini    : chr [1:3] "vodka" "espresso" "coffee liqueur"
##  $ zombie              : chr [1:6] "rum" "orange liqueur" "apricot brandy" "orange juice" ...
##  $ margarita           : chr [1:3] "tequila" "orange liqueur" "lime juice"
```

b) Create a tibble called `drinks`, where one column contains the name of the cocktail, and the other column contains the vector of ingredients

```
drinks = tibble(cocktail = names(cocktails),
                ingredients = cocktails)
```

c) Create a new column that contains the number of ingredients in each column using `mutate()` and **purrr**

```
drinks = drinks %>%
  mutate(total_ingredients = map(ingredients, length))
```

d) We're off out! Tonight we're particularly thirsty for a cocktail with rum in it. Filter `drinks` such that it only has cocktails containing rum

```
drinks %>%
  mutate(contains_rum = map_lgl(ingredients,
                                ~ any(.x == "rum"))) %>%
  filter(contains_rum)
## # A tibble: 3 x 4
##   cocktail             ingredients total_ingredients contains_rum
##   <chr>                <list>      <list>            <lgl>
## 1 pina_colada          <chr [3]>   <int [1]>         TRUE
## 2 long_island_iced_tea <chr [7]>   <int [1]>         TRUE
## 3 zombie               <chr [6]>   <int [1]>         TRUE
```

Question 2 - Beer !

So, we're at the pub with 8 mates and it's your round. In total you've been tasked with ordering 4 ales, 3 ipas, 1 stout plus an ale for yourself! We can load a data set of all of the ale, ipa and stouts that the pub sells from the course package

```
(data(beer_tidy, package = "jrTidyverse2"))
## [1] "beer_tidy"
```

We're going to randomly select each persons drink using **purrr**. If people had asked for an even number of ales, ipas and stouts we could have done this without **purrr** like so

```
beer_tidy %>%
  group_by(Type) %>%
  sample_n(3)
## # A tibble: 9 x 4
## # Groups:   Type [3]
##   URL                                ABV Color Type
##   <chr>                            <dbl> <dbl> <chr>
## 1 Gingered Ale                      4.22 17.3  Ale
## 2 Light House Cream Ale             6.9   8.09 Ale
## 3 Pale Pants Ale Mod                6    12.4  Ale
## 4 My Next Ipa                       7.64  7.17 Ipa
## 5 Out Of This World Ipa 15 Gallon   7.15 10.6  Ipa
## 6 07 Golden Spruce Ipa              5.77 10.3  Ipa
## 7 Tolmie Stout                      5.67 35.6  Stout
## 8 Vanilla Stout                     4.91 50    Stout
## 9 Oatmeal Stout                     6.13 41.4  Stout
```

a) Nest the data according to the drink `Type` and save it as `pub`.

```
pub = beer_tidy %>%
  nest(-Type)
## Warning: All elements of `...` must be named.
## Did you want `data = c(URL, ABV, Color)`?
```

b) Create a column called `n` that contains the total number of each drink `Type` you need to order

```
pub = pub %>%
  mutate(n = c(5, 3, 1))
```

c) Create a new column called order that contains the randomly sampled drinks you are going to order. You should be using `map2()` to parallel map over the columns `data` and `n`. You should also be using `sample_n()` to perform the sampling.

```
pub = pub %>%
  mutate(order = map2(data, n, sample_n))
# or
# mutate(order = map2(.x = data, .y = n, ~sample_n(.x, .y)))
```

d) To see the drinks, select only the `Type` and `order` column, then `unnest()`

```
pub %>%
  select(Type, order) %>%
  unnest()
## Warning: `cols` is now required.
## Please use `cols = c(order)`
## # A tibble: 9 x 4
```

```
##   Type  URL                                    ABV Color
##   <chr> <chr>                                <dbl> <dbl>
## 1 Ale   Firestone Double Barrel Ale Clone  4.85 14.0
## 2 Ale   461 Rainy Irish Red Ale            5.4  15.5
## 3 Ale   Hinterhof Belgian Pale Ale         4.87 14.9
## 4 Ale   Ss Red Ale                         5.32 14.9
## 5 Ale   Ejp Pumpkin Ale                    8.46 15
## 6 Ipa   Mosaic Ipa                         7.29 10.0
## 7 Ipa   Single Hop Sorachi Ace Double Ipa  8.95  6.39
## 8 Ipa   Jd S Ipa                           6.85  5
## 9 Stout Imperial Stout                     9.8  50
```

Question 3 - Happiness

You may remember the happiness data we used for practical 1 was recorded over 3 years; 2015, 2016 and 2017. For this question I've turned the happiness list in 3 tibbles, with each one representing the year. Running the following code will copy each file into your current working directory as a `.csv` file

```
library("jrTidyverse2")
get_happiness()
## Files have been copied successfully!
##           Check your current working directory.
```

a) Using a combination of **purrr** and the `unnest()` function from **tidyr**, read in and combine the 3 data sets. Don't delete the column containing the file name!

```
fnames = list.files("happiness", recursive = TRUE, full.names = TRUE)
happiness = tibble(fname = fnames) %>%
  mutate(data = map(fnames, read_csv)) %>%
  unnest()
```

b) The data within the csv files doesn't contain the year. Fortunately the file name does! Use the column containing the filename to create a column called `Year`. Have a look at the `str_remove()` or `parse_number()` functions from **stringr** and **readr** respectively

```
happiness = happiness %>%
  mutate(Year = parse_number(fname)) %>%
  select(-fname)
# or
# happiness = happiness %>%
#   mutate(Year = str_remove(fname, ".csv"),
#          Year = as.numeric(Year)) %>%
#   select(-fname)
```

c) Pick 3 countries and plot their happiness rank over time.

```
happiness %>%
  filter(Country %in% c("Denmark", "United Kingdom", "United States")) %>%
  ggplot(aes(x = Year, y = `Happiness Rank`, colour = Country)) +
  geom_line() +
  geom_point()
```

d) Every country in the data set has requested that they have their data sent to them individually. Make use of **purrr**'s parallel mapping functions and `nest()` to write the data to `.csv` files

```r
happiness_nest = happiness %>%
  nest(-Country)
map2(happiness_nest$Country, happiness_nest$data,
     ~write_csv(.y, path = paste0(.x, ".csv"))
```