Homework-3-Tidyverse

Task 1

Question a

We cannot directly use the read_csv() function because data.txt is delimited by semicolons and not commas. Instead, we can use read_csv2().

```r
library(tidyverse)
library(palmerpenguins)

data <- read_csv2("C:/Users/ZRose/Desktop/Skewl/ST_558/Homework3/data/data.txt")
data
```

```
# A tibble: 2 x 3
      x     y     z
  <dbl> <dbl> <dbl>
1     1     2     3
2     5     3     8
```

Question b

```r
data2 <- read_delim("C:/Users/ZRose/Desktop/Skewl/ST_558/Homework3/data/data2.txt",
                    delim = "6", col_types = "fdc")
data2
```

```
# A tibble: 3 x 3
  x         y z
  <fct> <dbl> <chr>
1 1         2 3
2 5         3 8
3 7         4 2
```

Task 2

Question a

```r
trailblazer <- read_csv(
  "C:/Users/ZRose/Desktop/Skewl/ST_558/Homework3/data/trailblazer.csv"
)
glimpse(trailblazer)
```

```
Rows: 9
Columns: 11
$ Player      <chr> "Damian Lillard", "CJ McCollum", "Norman Powell", "Robert ~
$ Game1_Home  <dbl> 20, 24, 14, 8, 20, 5, 11, 2, 7
$ Game2_Home  <dbl> 19, 28, 16, 6, 9, 5, 18, 8, 11
$ Game3_Away  <dbl> 12, 20, NA, 0, 4, 8, 12, 5, 5
$ Game4_Home  <dbl> 20, 25, NA, 3, 17, 10, 17, 8, 9
$ Game5_Home  <dbl> 25, 14, 12, 9, 14, 9, 5, 3, 8
$ Game6_Away  <dbl> 14, 25, 14, 6, 13, 6, 19, 8, 8
$ Game7_Away  <dbl> 20, 20, 22, 0, 7, 0, 17, 7, 4
$ Game8_Away  <dbl> 26, 21, 23, 6, 6, 7, 15, 0, 0
$ Game9_Home  <dbl> 4, 27, 25, 19, 10, 0, 16, 2, 7
$ Game10_Home <dbl> 25, 7, 13, 12, 15, 6, 10, 4, 8
```

Question b

```r
trailblazer_longer <- trailblazer |>
  pivot_longer(
    cols = starts_with("Game"),
    names_to = "GameLabel",
    values_to = "Points"
  ) |>
  separate_wider_delim(
    cols = GameLabel,
    delim = "_",
    names = c("Game", "Location")
  )

head(trailblazer_longer, 5)
```

```
# A tibble: 5 x 4
  Player         Game  Location Points
  <chr>          <chr> <chr>     <dbl>
1 Damian Lillard Game1 Home         20
2 Damian Lillard Game2 Home         19
3 Damian Lillard Game3 Away         12
4 Damian Lillard Game4 Home         20
5 Damian Lillard Game5 Home         25
```

Question c

```
trailblazer_longer |>
  pivot_wider(
    names_from = Location,
    values_from = Points
  ) |>
  group_by(Player) |>
  summarise(
    mean_home = mean(Home, na.rm = TRUE),
    mean_away = mean(Away, na.rm = TRUE),
    diff_home_away = mean_home - mean_away
  ) |>
  arrange(desc(diff_home_away))
```

```
# A tibble: 9 x 4
  Player           mean_home mean_away diff_home_away
  <chr>                <dbl>     <dbl>          <dbl>
1 Jusuf Nurkic          14.2       7.5           6.67
2 Robert Covington       9.5       3             6.5
3 Nassir Little          8.33      4.25          4.08
4 Damian Lillard        18.8      18             0.833
5 Cody Zeller            5.83      5.25          0.583
6 Larry Nance Jr         4.5       5            -0.5
7 CJ McCollum           20.8      21.5          -0.667
8 Anfernee Simons       12.8      15.8          -2.92
9 Norman Powell         16        19.7          -3.67
```

On average, Jusuf Nurkic, Robert Covington, Nassir Little, Damian Lillard, and Cody Zeller scored more points at home through the first 10 games of the season than away.

Task 3

Question a

Species and island aren't unique identifiers for each row. This is because there can be multiple penguins of the same species on each island. Some entries are $< NULL >$ because there aren't any values for that combination of species and island. Some values are like $< dbl\ [52] >$ because there are many values for that combination of species and island (52 of them). The way these combinations are selected results in a list of doubles. We see that the columns are $< list >$ as they contain a list of bill length values rather than a single value.

Question b

```
penguins |>
  group_by(species, island) |>
  summarise(n = as.double(n())) |>
  pivot_wider(names_from = island, values_from = n, values_fill = 0)
```

`summarise()` has grouped output by 'species'. You can override using the
`.groups` argument.

```
# A tibble: 3 x 4
# Groups:   species [3]
  species   Biscoe Dream Torgersen
  <fct>      <dbl> <dbl>     <dbl>
1 Adelie        44    56        52
2 Chinstrap      0    68         0
3 Gentoo       124     0         0
```

Task 4

```
penguins |>
  mutate(bill_length_mm = case_when(
    is.na(bill_length_mm) & species == "Adelie" ~ 26,
    is.na(bill_length_mm) & species == "Gentoo" ~ 30,
    TRUE ~ bill_length_mm
  )) |>
  arrange(bill_length_mm) |>
  slice_head(n = 10)
```

```
# A tibble: 10 x 8
   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
   <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
 1 Adelie  Torgersen           26            NA                NA          NA
 2 Gentoo  Biscoe              30            NA                NA          NA
 3 Adelie  Dream               32.1          15.5             188        3050
 4 Adelie  Dream               33.1          16.1             178        2900
 5 Adelie  Torgersen           33.5          19               190        3600
 6 Adelie  Dream               34            17.1             185        3400
 7 Adelie  Torgersen           34.1          18.1             193        3475
 8 Adelie  Torgersen           34.4          18.4             184        3325
 9 Adelie  Biscoe              34.5          18.1             187        2900
10 Adelie  Torgersen           34.6          21.1             198        4400
# i 2 more variables: sex <fct>, year <int>
```