Assignment 3

Kevin Ha - 571821   Ola Andre Olofsson - 170745

Answer the following questions

1. What information does the file ddf_concepts.csv contain? Den inneholder blant annet "description" av diverse variabler.

2. What information does the file ddf–entities–geo–country.csv contain? Den tilskriver land ulike faktorer som:

- Innenlandsstat eller stat med kystlinje
- Geografiske koordinater med lengde- og breddegrader
- Hovedreligion
- Medlemskap i FN

3. What information does the file ddf–entities–geo–un_sdg_region.csv contain? Den inneholder informasjon om åtte ulike regioner som er FN-regioner. Dette fremkommer delvis av filnavnet som inneholder UN. SDG tenker jeg står for Sustainable Development Goals, men ser ikke helt sammenhengen her ennå.

4. What variables does the gapminder dataset from the gapminder package contain? To what continent are Australia and New Zealand assigned?

```
library(gapminder)
View(gapminder)
```

Australia og New Zealing er angitt til Oseania. Datasettet viser landsliste med kontinentnavn, forventet levealder, BNP og populasjon.

5. Recreate the continent variable in gapminder with the new data. Only include countries that have a iso3166_1_alpha3 code. Use data from ddf–entities–geo–country.csv and call this tibble g_c. Let g_c be your main tibble in the following, i.e. add variables to this tibble.

```
library(readr)
g_c <- read_csv("data/ddf--entities--geo--country.csv")
```

```
## Rows: 273 Columns: 22
```

```
## -- Column specification -----------------------------------------------------
## Delimiter: ","
## chr (17): country, g77_and_oecd_countries, income_3groups, income_groups, is...
## dbl  (3): iso3166_1_numeric, latitude, longitude
## lgl  (2): is--country, un_state
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
as_tibble(g_c)
```

```
## # A tibble: 273 x 22
##    country   g77_and_oecd_countries income_3groups income_groups   'is--country'
##    <chr>     <chr>                  <chr>          <chr>           <lgl>
##  1 abkh      others                 <NA>           <NA>            TRUE
##  2 abw       others                 high_income    high_income     TRUE
##  3 afg       g77                    low_income     low_income      TRUE
##  4 ago       g77                    middle_income  lower_middle_i~ TRUE
##  5 aia       others                 <NA>           <NA>            TRUE
##  6 akr_a_dhe others                 <NA>           <NA>            TRUE
##  7 ala       others                 <NA>           <NA>            TRUE
##  8 alb       others                 middle_income  upper_middle_i~ TRUE
##  9 and       others                 high_income    high_income     TRUE
## 10 ant       others                 <NA>           <NA>            TRUE
## # ... with 263 more rows, and 17 more variables: iso3166_1_alpha2 <chr>,
## #   iso3166_1_alpha3 <chr>, iso3166_1_numeric <dbl>, iso3166_2 <chr>,
## #   landlocked <chr>, latitude <dbl>, longitude <dbl>,
## #   main_religion_2008 <chr>, name <chr>, un_sdg_ldc <chr>,
## #   un_sdg_region <chr>, un_state <lgl>, unhcr_region <chr>,
## #   unicef_region <chr>, unicode_region_subtag <chr>, world_4region <chr>,
## #   world_6region <chr>
```

```
# Angir ønsket navn, g_c, til datasettet

g_c <- g_c %>%
```

```
# Lager ny variabel med case_when-funksjonen som lar oss vektorisere flere vilkår (sta
  mutate(continent = case_when(
    world_4region == "asia" & un_sdg_region %in% c(
      "un_australia_and_new_zealand",
      "un_oceania_exc_australia_and_new_zealand") ~ "Oceania",

    world_4region ==  "asia" & !(un_sdg_region %in% c(
      "un_australia_and_new_zealand",
      "un_oceania_exc_australia_and_new_zealand")) ~ "Asia",

    world_4region == "africa" ~ "Africa",
    world_4region == "americas" ~ "Americas",
    world_4region ==  "europe" ~ "Europe")
  ) %>%
 filter(!is.na(iso3166_1_alpha3))
```

6. How many countries are there now?

```
# Teller antall rader med land, og vi får 247 land
nrow(g_c)
```

```
## [1] 247
```

```
# Alternativt kan vi bruke length(unique)
length(unique(g_c$country))
```

```
## [1] 247
```

6 / 7 ?. How many countries are there now in each continent?

```
g_c %>%
  group_by(continent) %>%
  summarise(countries = length(unique(country)))
```

```
## # A tibble: 5 x 2
##   continent countries
##   <chr>         <int>
## 1 Africa           59
## 2 Americas         55
## 3 Asia             47
## 4 Europe           58
## 5 Oceania          28
```