R Take-Home Assignment 1: Gapminder (Wrangling & Visualization)

Your Name

2025-09-30

Overview

In this assignment you will work with the **Gapminder dataset**, which contains information on life expectancy, population size, and GDP per capita for 142 countries between 1952 and 2007.
You will practice importing data, exploring its structure, subsetting, summarising, and producing plots to investigate global health and economic patterns.

- https://github.com/resbaz/r-novice-gapminder-files/raw/master/data/gapminder-FiveYearData.csv
- **Skills:** loading CSVs, subsetting, summarising, plotting
- **Deliverable:** knitted **HTML** plus your completed `.Rmd`

Preparation

You will need the following R packages:

- `dplyr`
- `ggplot2`
- `readr`

```
library(dplyr)
library(ggplot2)
library(readr)
```

1. Load the data

The starting point of any analysis is importing the dataset. Here, you will use the URL provided above.

Task: Load the Gapminder dataset into R and store it as `gapminder`.

```
url <- "https://github.com/resbaz/r-novice-gapminder-files/raw/master/data/gapminder-FiveYearData.csv"
gapminder <- read_csv(url, show_col_types = FALSE)
head(gapminder)
```

```
## # A tibble: 6 x 6
##   country      year      pop continent lifeExp gdpPercap
##   <chr>       <dbl>    <dbl> <chr>       <dbl>     <dbl>
## 1 Afghanistan  1952  8425333 Asia         28.8      779.
```

```
## 2 Afghanistan  1957  9240934 Asia         30.3      821.
## 3 Afghanistan  1962 10267083 Asia         32.0      853.
## 4 Afghanistan  1967 11537966 Asia         34.0      836.
## 5 Afghanistan  1972 13079460 Asia         36.1      740.
## 6 Afghanistan  1977 14880372 Asia         38.4      786.
```

2. Explore the dataset

Before doing analysis, it is important to understand the dataset's scope. You should examine how many countries it contains, which years are covered, and what variables are available.

Task: Report the number of countries, the range of years, and the variable names.

```
n_distinct(gapminder$country)
```

```
## [1] 142
```

```
range(gapminder$year)
```

```
## [1] 1952 2007
```

```
names(gapminder)
```

```
## [1] "country"   "year"      "pop"       "continent" "lifeExp"   "gdpPercap"
```

Write your answers here:

```
Countries: ...
Years: ...
Variables: ...
```

3. Subset the data

Analyses are often focused on a specific region. For this step, you will work only with European countries.

Task: Create a subset containing only the rows where `continent` is `"Europe"`. Save this as `europe`.

```
europe <- gapminder %>% filter(continent == "Europe")
head(europe)
```

```
## # A tibble: 6 x 6
##   country  year     pop continent lifeExp gdpPercap
##   <chr>   <dbl>   <dbl> <chr>       <dbl>     <dbl>
## 1 Albania  1952 1282697 Europe       55.2     1601.
## 2 Albania  1957 1476505 Europe       59.3     1942.
## 3 Albania  1962 1728137 Europe       64.8     2313.
## 4 Albania  1967 1984060 Europe       66.2     2760.
## 5 Albania  1972 2263554 Europe       67.7     3313.
## 6 Albania  1977 2509048 Europe       68.9     3533.
```

4. Summarise life expectancy

To see general patterns across the world, it is helpful to calculate average values. By summarising life expectancy by continent and year, you can compare how regions developed over time.

Task: Calculate the mean life expectancy per continent per year. Save the result as `lifeexp_summary`.

```r
lifeexp_summary <- gapminder %>%
  group_by(continent, year) %>%
  summarise(mean_lifeExp = mean(lifeExp), .groups = "drop")
head(lifeexp_summary)
```

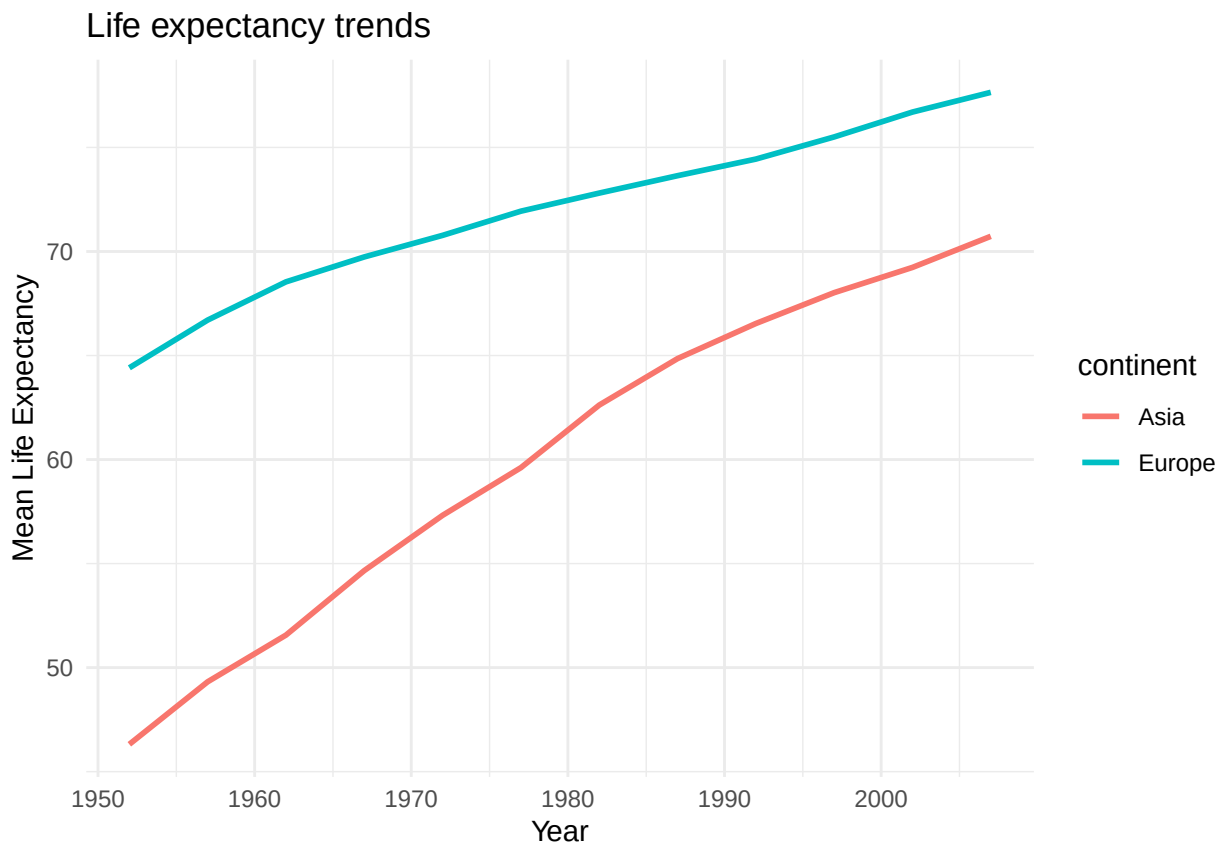
```
## # A tibble: 6 x 3
##   continent  year mean_lifeExp
##   <chr>     <dbl>        <dbl>
## 1 Africa     1952         39.1
## 2 Africa     1957         41.3
## 3 Africa     1962         43.3
## 4 Africa     1967         45.3
## 5 Africa     1972         47.5
## 6 Africa     1977         49.6
```

5. Visualise trends

Summaries are useful, but plots reveal patterns more clearly. A line plot of life expectancy over time allows you to compare trajectories between regions.

Task: Produce a line plot of life expectancy over time for **two continents of your choice**. Label the axes and add a title.

```r
lifeexp_summary %>% filter(continent %in% c("Europe", "Asia")) %>%
  ggplot(aes(year, mean_lifeExp, color=continent)) +
  geom_line(linewidth=1) +
  labs(title="Life expectancy trends", x="Year", y="Mean Life Expectancy") +
  theme_minimal()
```



6. Reflection

Numbers and plots show differences, but interpretation is just as important. Take a moment to reflect on what you observe.

Task: Write 3–4 sentences describing the patterns you see across continents in life expectancy trends.

Write your reflection here...

Appendix

```
sessionInfo()
```

```
## R version 4.5.1 (2025-06-13)
## Platform: aarch64-apple-darwin24.4.0
## Running under: macOS Sequoia 15.6.1
##
## Matrix products: default
## BLAS:   /opt/homebrew/Cellar/openblas/0.3.30/lib/libopenblasp-r0.3.30.dylib
## LAPACK: /opt/homebrew/Cellar/r/4.5.1/lib/R/lib/libRlapack.dylib;  LAPACK version 3.12.1
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: Europe/Amsterdam
```

```
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] readr_2.1.5   ggplot2_3.5.2 dplyr_1.1.4
##
## loaded via a namespace (and not attached):
##  [1] crayon_1.5.3       vctrs_0.6.5        cli_3.6.5          knitr_1.50
##  [5] rlang_1.1.6        xfun_0.52          generics_0.1.4     labeling_0.4.3
##  [9] bit_4.6.0          glue_1.8.0         htmltools_0.5.8.1  hms_1.1.3
## [13] scales_1.4.0       rmarkdown_2.30     grid_4.5.1         evaluate_1.0.4
## [17] tibble_3.2.1       tzdb_0.5.0         fastmap_1.2.0      yaml_2.3.10
## [21] lifecycle_1.0.4    compiler_4.5.1     RColorBrewer_1.1-3 pkgconfig_2.0.3
## [25] farver_2.1.2       digest_0.6.37      R6_2.6.1           utf8_1.2.4
## [29] tidyselect_1.2.1   curl_6.2.2         parallel_4.5.1     vroom_1.6.5
## [33] pillar_1.10.2      magrittr_2.0.3     bit64_4.6.0-1      tools_4.5.1
## [37] withr_3.0.2        gtable_0.3.6
```