Math 300 NTI Lesson 9

Case Study and Review

Professor Bradley Warner

June, 2022

Contents

Objectives

1. Read data into `R` and convert it to tidy format.

Reading

Chapter 4.3 - 4.5

Lesson

Remember that you will be running this more like a lab than a lecture. You want them using `R` and answering questions. Have them open the notes rmd and work through it together.

Work through the learning checks LC4.4 - LC4.5.

- You can use the GUI to get the code to read the Excel file from the web. Then post the code into your rmarkdown file. Remember, to knit the file it has to read the data into the temporary workspace.
- Save time to review for the GR.

Setup

```
library(dplyr)
library(ggplot2)
library(readr)
library(tidyr)
library(nycflights13)
library(fivethirtyeight)
```

LC 4.4 (Objective 1)

(LC4.4) Convert the `dem_score` data frame into a tidy data frame and assign the name of `dem_score_tidy` to the resulting long-formatted data frame.

Solution:

```
# Get the data
library(readxl)
url <- "https://moderndive.com/data/dem_score.xlsx"
destfile <- "dem_score.xlsx"
curl::curl_download(url, destfile)
dem_score <- read_excel(destfile)
```

```
head(dem_score)
```

```
## # A tibble: 6 x 10
##   country    `1952` `1957` `1962` `1967` `1972` `1977` `1982` `1987` `1992`
##   <chr>       <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Albania        -9     -9     -9     -9     -9     -9     -9     -9      5
## 2 Argentina      -9     -1     -1     -9     -9     -9     -8      8      7
## 3 Armenia        -9     -7     -7     -7     -7     -7     -7     -7      7
## 4 Australia      10     10     10     10     10     10     10     10     10
## 5 Austria        10     10     10     10     10     10     10     10     10
## 6 Azerbaijan     -9     -7     -7     -7     -7     -7     -7     -7      1
```

```
dem_score_tidy <- dem_score %>%
  pivot_longer(
    names_to = "year", values_to = "democracy_score",
    cols = -country
  )
```

```
head(dem_score_tidy)
```

```
## # A tibble: 6 x 3
##   country year  democracy_score
##   <chr>   <chr>           <dbl>
## 1 Albania 1952               -9
## 2 Albania 1957               -9
## 3 Albania 1962               -9
## 4 Albania 1967               -9
## 5 Albania 1972               -9
## 6 Albania 1977               -9
```

Let's now compare the `dem_score` and `dem_score_tidy`. `dem_score` has democracy score information for each year in columns, whereas in `dem_score_tidy` there are explicit variables `year` and `democracy_score`. While both representations of the data contain the same information, it is easier to use `ggplot()` to create plots using the `dem_score_tidy` data frame.

LC 4.5 (Objective 1)

(LC4.5) Read in the life expectancy data stored at https://moderndive.com/data/le_mess.csv and convert it to a tidy data frame.

Solution: The code is similar

```
life_expectancy <- read_csv("https://moderndive.com/data/le_mess.csv")
```

```
head(life_expectancy)
```

```
## # A tibble: 6 x 67
##   country  `1951` `1952` `1953` `1954` `1955` `1956` `1957` `1958` `1959` `1960`
##   <chr>     <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Afghani~   27.1   27.7   28.2   28.7   29.3   29.8   30.3   30.9   31.4   31.9
## 2 Albania    54.7   55.2   55.8   56.6   57.4   58.4   59.5   60.6   61.8   62.9
## 3 Algeria    43.0   43.5   44.0   44.4   44.9   45.4   45.9   46.4   47.0   47.5
## 4 Angola     31.0   31.6   32.1   32.7   33.2   33.8   34.3   34.9   35.4   36.0
## 5 Antigua~   58.3   58.8   59.3   59.9   60.4   60.9   61.4   62.0   62.5   63.0
## 6 Argenti~   61.9   62.5   63.1   63.6   64.0   64.4   64.7   65     65.2   65.4
## # ... with 56 more variables: `1961` <dbl>, `1962` <dbl>, `1963` <dbl>,
## #   `1964` <dbl>, `1965` <dbl>, `1966` <dbl>, `1967` <dbl>, `1968` <dbl>,
## #   `1969` <dbl>, `1970` <dbl>, `1971` <dbl>, `1972` <dbl>, `1973` <dbl>,
## #   `1974` <dbl>, `1975` <dbl>, `1976` <dbl>, `1977` <dbl>, `1978` <dbl>,
## #   `1979` <dbl>, `1980` <dbl>, `1981` <dbl>, `1982` <dbl>, `1983` <dbl>,
## #   `1984` <dbl>, `1985` <dbl>, `1986` <dbl>, `1987` <dbl>, `1988` <dbl>,
## #   `1989` <dbl>, `1990` <dbl>, `1991` <dbl>, `1992` <dbl>, `1993` <dbl>, ...
```

```
life_expectancy_tidy <- life_expectancy %>%
  pivot_longer(
    names_to = "year",
    values_to = "life_expectancy",
    cols = -country
  )
```

We observe the same construct structure with respect to `year` in `life_expectancy` vs `life_expectancy_tidy` as we did in `dem_score` vs `dem_score_tidy`:

```
head(life_expectancy_tidy)
```

```
## # A tibble: 6 x 3
##   country     year  life_expectancy
##   <chr>       <chr>           <dbl>
## 1 Afghanistan 1951             27.1
## 2 Afghanistan 1952             27.7
## 3 Afghanistan 1953             28.2
## 4 Afghanistan 1954             28.7
## 5 Afghanistan 1955             29.3
## 6 Afghanistan 1956             29.8
```

Documenting software

- File creation date: 2022-06-16
- R version 4.1.3 (2022-03-10)
- `ggplot2` package version: 3.3.6
- `tidyr` package version: 1.2.0

- `readr` package version: 2.1.2

- `dplyr` package version: 1.0.9
- `nycflights13` package version: 1.0.2
- `fivethirtyeight` package version: 0.6.2