WK8_HW

Lidan Gao

Homework

Due: March 21st, 2024 before 3:00 pm

1. How can you check for missing values (NA) in the **fert_cons_data** dataset? Please provide the R code and give a brief explanation of what is happening in the code.

   I can use **summary()** to check NA. summary() can return a summary of the dataset, including the min, max, median, mean and NA's.

```
# summary(fert_cons_data)
```

2. What is the purpose of reshaping the **fert_cons_data** dataset into a wide format?

   Reshaping a dataset into a wide format can make it easier to analyze data by subject (e.g., country) across time or categories since each row represents a subject and each column represents a time period or category. It's useful for running certain types of analyses, creating specific visualizations, or exporting data for reports where a wide format is more interpretable.

3. How can you rename the columns **Year** and **Fert** in the **fert_long** dataset? Please provide the R code and give a brief explanation of what is happening in the code.

   I use dplyr's function rename() to rename the columns **Year** and **Fert** to **year** and **fert_cons** in the **fert_long** dataset**.**

```
# fert_long <- dplyr::rename(fert_long,
#                            year = Year,
#                            fert_cons = Fert)
```

4. What function is used to create a density plot of the **fert_cons** variable in the **fert_long** dataset? Please provide the R code and give a brief explanation of what is happening in the code.

   I use ggplot2's function ggplot and geom_density() to create a density plot for the fert_long dataset. X axis is fert_cons.

```
#library(ggplot2)

#ggplot(data = fert_long, aes(x=fert_cons)) +
#geom_density() +
#theme_bw()
```

5. How can you recode the country name "Korea, Rep." to "South Korea" in the **fert_long_sub** dataset? Please provide the R code and give a brief explanation of what is happening in the code.

   I use Recode country == "Korea, Rep." to "South Korea" to change the country name.

```
# Recode country == "Korea, Rep." to "South Korea"

# fert_long_sub$country[fert_long_sub$country ==
# "Korea, Rep."] <- "South Korea"

# fert_long_sub
```

6. How is the **fert_cons_log** variable created in the **fert_long_sub** dataset? Please provide the R code and give a brief explanation of what is happening in the code.

   I use log() to change the variable name.

```
# use log() to change the variable name
# fert_long_sub$fert_cons_log <- log(fert_long_sub$fert_cons)
```

7. What is the purpose of creating the **fert_cons_group** variable in the **fert_long_sub** dataset?

   Creating the **fert_cons_group** variable can help in categorizing continuous data into discrete groups for easier analysis or visualization. It's useful for summarizing data, performing group-based analyses, and making the data more interpretable.

8. How can you convert the **fert_cons_group** variable into a factor variable with labels "low", "medium low", "medium high", and "high"? Please provide the R code and give a brief explanation of what is happening in the code.

First I defined the fc_labels by using **c()** and then convert the **fert_cons_group** variable into a factor variable by using **factor().**

```
# Create vector of factor level labels

# fc_labels <- c("low", "medium low", "medium high", "high")

# Convert fert_cons_group to a factor

#fert_long_sub$fert_cons_group <- factor(fert_long_sub$fert_cons_group, labels = fc_l
```

9. What is the purpose of using the **countrycode** package in the **iso2c** variable creation?

   The **countrycode** package is used to convert country names and codes between different naming conventions. For **iso2c** variable creation, it can standardize country identifiers, making it easier to merge datasets from different sources that may use different conventions for country names or codes.

10. How can you merge the **fin_regulator**, **disprop_data**, and **fert_long_sub** datasets based on the **iso2c** variable? Please provide the R code and give a brief explanation of what is happening in the code.

I merged fin_regulator and disprop_data by **iso2c** first and then merged the combined data set with and fert_long_sub by **iso2c.**

```
# Merge fin_regulator and disprop_data

#merged_data_1 <- merge(x = fin_regulator, y = disprop_data, by = "iso2c", all = TRUE)

# Merge combined data set with and fert_long_sub

#merged_data_1 <- merge(x = merged_data_1, y = fert_long_sub, by = "iso2c", all = TRUE)
```

Turn in a PDF. On your github account Push, (1) PDF, (2) Quarto document, and (3) .README file explaining what analysis, software, was used to a new folder. Include a link to your Github repository in the PDF.

https://github.com/ldgao11/Chapter_7.git