Quiz 2 preparation solution

Answer the following questions.

These questions relate to the homework problem 12.2.1 Exercise 2. Use what you know from Chapter 13 to answer the questions for *table4a* and *table4b*.

1. Using *table4a* and *table4b* from the *tidyverse* R package, merge the two dataframes into one and create a column, rate per 10000 variable, for each year.

library(tidyverse)  
table4a

## # A tibble: 3 x 3  
## country `1999` `2000`  
## \* <chr> <int> <int>  
## 1 Afghanistan 745 2666  
## 2 Brazil 37737 80488  
## 3 China 212258 213766

table4b

## # A tibble: 3 x 3  
## country `1999` `2000`  
## \* <chr> <int> <int>  
## 1 Afghanistan 19987071 20595360  
## 2 Brazil 172006362 174504898  
## 3 China 1272915272 1280428583

table\_new2 <- table4a %>% inner\_join(table4b, by = c("country"))  
table\_new2

## # A tibble: 3 x 5  
## country `1999.x` `2000.x` `1999.y` `2000.y`  
## <chr> <int> <int> <int> <int>  
## 1 Afghanistan 745 2666 19987071 20595360  
## 2 Brazil 37737 80488 172006362 174504898  
## 3 China 212258 213766 1272915272 1280428583

table\_new2a <- table\_new2 %>% mutate(  
rate.1999 = (`1999.x`/`1999.y`)\*10000,  
rate.2000 = (`2000.x`/`2000.y`)\*10000  
) %>%  
select(country, rate.1999, rate.2000)  
table\_new2a

## # A tibble: 3 x 3  
## country rate.1999 rate.2000  
## <chr> <dbl> <dbl>  
## 1 Afghanistan 0.373 1.29  
## 2 Brazil 2.19 4.61  
## 3 China 1.67 1.67

1. Is your final dataframe for questions 1 tidy? Yes or no, explain.

**Answer:** No. The *table\_new2a* is not tidy. The rates are in two columns.

1. Convert your final dataframe for question 1 into a tidy dataframe with three columns country, year and rate.

**Answer:** Gather the columns into year and rate columns.

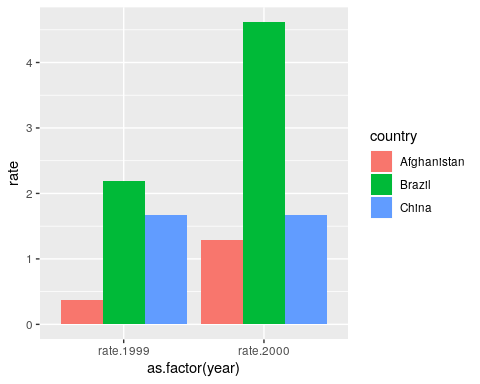
table\_new2a %>% gather(rate.1999, rate.2000, key = "year", value = "rate")

## # A tibble: 6 x 3  
## country year rate  
## <chr> <chr> <dbl>  
## 1 Afghanistan rate.1999 0.373  
## 2 Brazil rate.1999 2.19   
## 3 China rate.1999 1.67   
## 4 Afghanistan rate.2000 1.29   
## 5 Brazil rate.2000 4.61   
## 6 China rate.2000 1.67

1. Make a clustered bar graph displaying the data.

Note the use of the as.factor() function.

table\_new2a %>% gather(rate.1999, rate.2000, key = "year", value = "rate") %>%  
 ggplot(aes(x = as.factor(year), y = rate, fill = country) ) +  
 geom\_bar(position="dodge", stat="identity")



The next question relates to Chapter 14.

library(stringr)

1. For the string “Today is the second quiz.”
   1. Use an *str\_?* R function to count the length of the string.
   2. Use an *str\_* R function to change all of the letters to lower case.
   3. Use an *str\_?* R function to subset the string into separate words.

x <- "Today is the second quiz."  
x

## [1] "Today is the second quiz."

# a.  
str\_count(x)

## [1] 25

# b.  
str\_to\_lower(x)

## [1] "today is the second quiz."

# c.  
str\_split(x, " ")

## [[1]]  
## [1] "Today" "is" "the" "second" "quiz."