Chaudhary_homework_7

Jyoti Chaudhary

November 6, 2016

TASK 1

Download the csv file on maternal mortality at http://apps.who.int/gho/athena/data/xmart.csv?target=GHO/MDG_0000000025,MDG_0000000026&profile=crosstable&filter=COUNTRY:*;YEAR:*&x-sideaxis=COUNTRY;YEAR&x-topaxis=GHO (http://apps.who.int/gho/athena/data/xmart.csv?target=GHO/MDG_0000000025,MDG_0000000026&profile=crosstable&filter=COUNTRY:*;YEAR:*&x-sideaxis=COUNTRY;YEAR&x-topaxis=GHO) and store it in a data directory in your project repo. (You can read about the data at http://apps.who.int/gho/data/view.main.SDG31v?lang=en (http://apps.who.int/gho/data/view.main.SDG31v?lang=en)) .Although the data is tidy, it needs cleaning. Notice that spaces are used instead of commas in numbers. Turn the numbers for maternal mortality inside square brackets into two new variables, lower_bound and upper_bound

```
xmart <- read_csv("xmart.csv")

colnames(xmart) <- c("Country", "Year", "Mortality.ratio", "Birth.attended")

xmart$Year <- str_sub(xmart$Year, end = 4)

xmart$pos_openbraces <- str_locate(xmart$Mortality.ratio, "\\[")
xmart$pos_closebraces <- str_locate(xmart$Mortality.ratio, "\\]")
xmart$pos_dash <- str_locate(xmart$Mortality.ratio, "-")


xmart$pos_openbraces <- xmart$pos_openbraces[,1]
xmart$pos_closebraces <- xmart$pos_closebraces[,1]
xmart$pos_dash <- xmart$pos_dash[,1]


xmart$lower_bound <- ifelse(xmart$pos_openbraces != "NA", str_sub(xmart$Mortality.ratio, xmart$p
os_openbraces + 1, xmart$pos_dash - 1), "NA")

xmart$upper_bound <- ifelse(xmart$pos_openbraces != "NA", str_sub(xmart$Mortality.ratio, xmart$p
os_dash + 1, xmart$pos_closebraces -1), "NA")

xmart$Mortality.ratio <- ifelse(xmart$pos_openbraces != "NA", str_sub(xmart$Mortality.ratio, end
 = xmart$pos_openbraces - 1), "NA")

xmart$Mortality.ratio <- as.integer(str_replace_all(xmart$Mortality.ratio, " ", ""))
xmart$lower_bound <- as.integer(str_replace_all(xmart$lower_bound, " ", ""))
xmart$upper_bound <- as.integer(str_replace_all(xmart$upper_bound, " ", ""))

xmart_clean <- select(xmart, Country, Year, Mortality.ratio, Birth.attended, lower_bound, upper_
bound)

head(xmart_clean)
```

```
## # A tibble: 6 × 6
##       Country  Year Mortality.ratio Birth.attended lower_bound upper_bound
##         <chr> <chr>           <int>          <dbl>       <int>       <int>
## 1 Afghanistan  2015             396             NA         253         620
## 2 Afghanistan  2013              NA           45.2          NA          NA
## 3 Afghanistan  2000            1100             NA         745        1570
## 4 Afghanistan  1990            1340             NA         878        1950
## 5     Albania  2015              29             NA          16          46
## 6     Albania  2008              NA           99.3          NA          NA
```

TASK 2

Download the gross national income data from http://apps.who.int/gho/data/node.main.GNI107?lang=en (http://apps.who.int/gho/data/node.main.GNI107?lang=en) (Download the complete data as a “CVS table”, storing it in the data folder). Tidy this data, so that there is one entry per country per year

```
income_data <- read_csv("data.csv", col_names = TRUE, col_types = NULL, skip = 1)
tall_income_data <- gather(income_data, key="Year", value = "Income", 2:25)
tall_income_data$Income <-  as.integer(str_replace_all(tall_income_data$Income, " ", ""))
clean_income_data <- filter(tall_income_data, tall_income_data$Income != "NA")

# Remove duplicate rows on Country and Year
clean_income_data <- clean_income_data[!duplicated(clean_income_data[,c('Country', 'Year')]),]

colnames(clean_income_data) <- c("COUNTRY", "YEAR", "INCOME")

head(clean_income_data)
```

```
## # A tibble: 6 × 3
##               COUNTRY  YEAR INCOME
##                 <chr> <chr>  <int>
## 1         Afghanistan  2013   2000
## 2             Albania  2013  10520
## 3             Algeria  2013  12990
## 4              Angola  2013   6770
## 5 Antigua and Barbuda  2013  20070
## 6             Armenia  2013   8140
```

TASK 3

Merge the two files by country and year, retaining only rows for which you have per capita income and either maternal mortality or percentage attended births

```
merge_data <- merge(clean_income_data, xmart_clean, by.x = c("COUNTRY", "YEAR"), by.y = c("Count
ry", "Year")) %>% filter(INCOME != "NA" & (Mortality.ratio != "NA" | Birth.attended != "NA"))

head(merge_data)
```

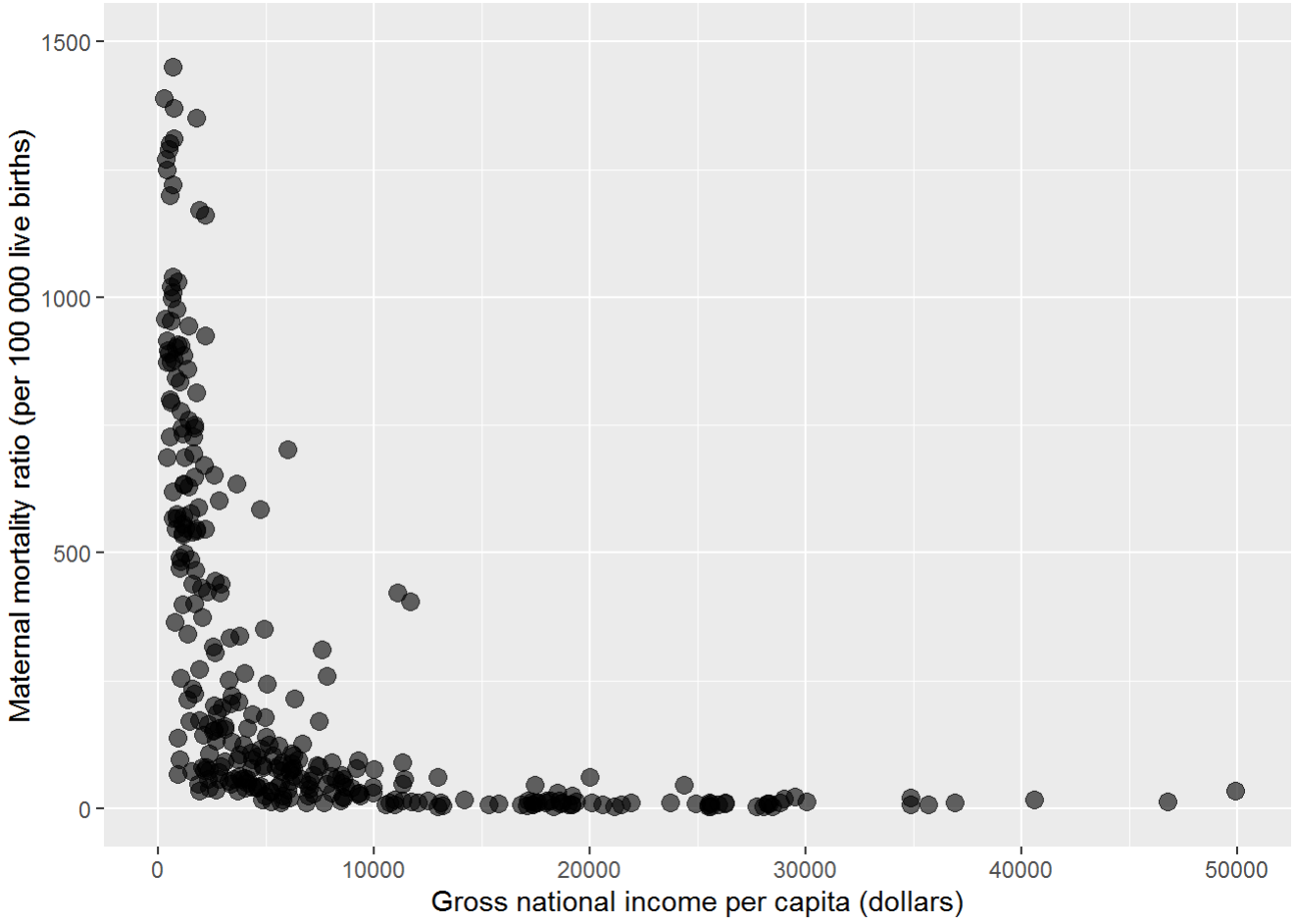
```
##       COUNTRY YEAR INCOME Mortality.ratio Birth.attended lower_bound
## 1 Afghanistan 2013   2000              NA           45.2          NA
## 2     Albania 1990   2840              71             NA          58
## 3     Albania 2000   4370              43             NA          33
## 4     Albania 2008   8920              NA           99.3          NA
## 5     Algeria 1990   6330             216             NA         141
## 6     Algeria 2000   7460             170             NA         118
##   upper_bound
## 1          NA
## 2          88
## 3          56
## 4          NA
## 5         327
## 6         241
```

TASK 4

Make a two polished and informative graphs, one relating income to maternal mortality and the other relating income to percentage attended births Save a csv file with the merged, cleaned, and tidy data, using a suitable name of your choosing in the data folder.

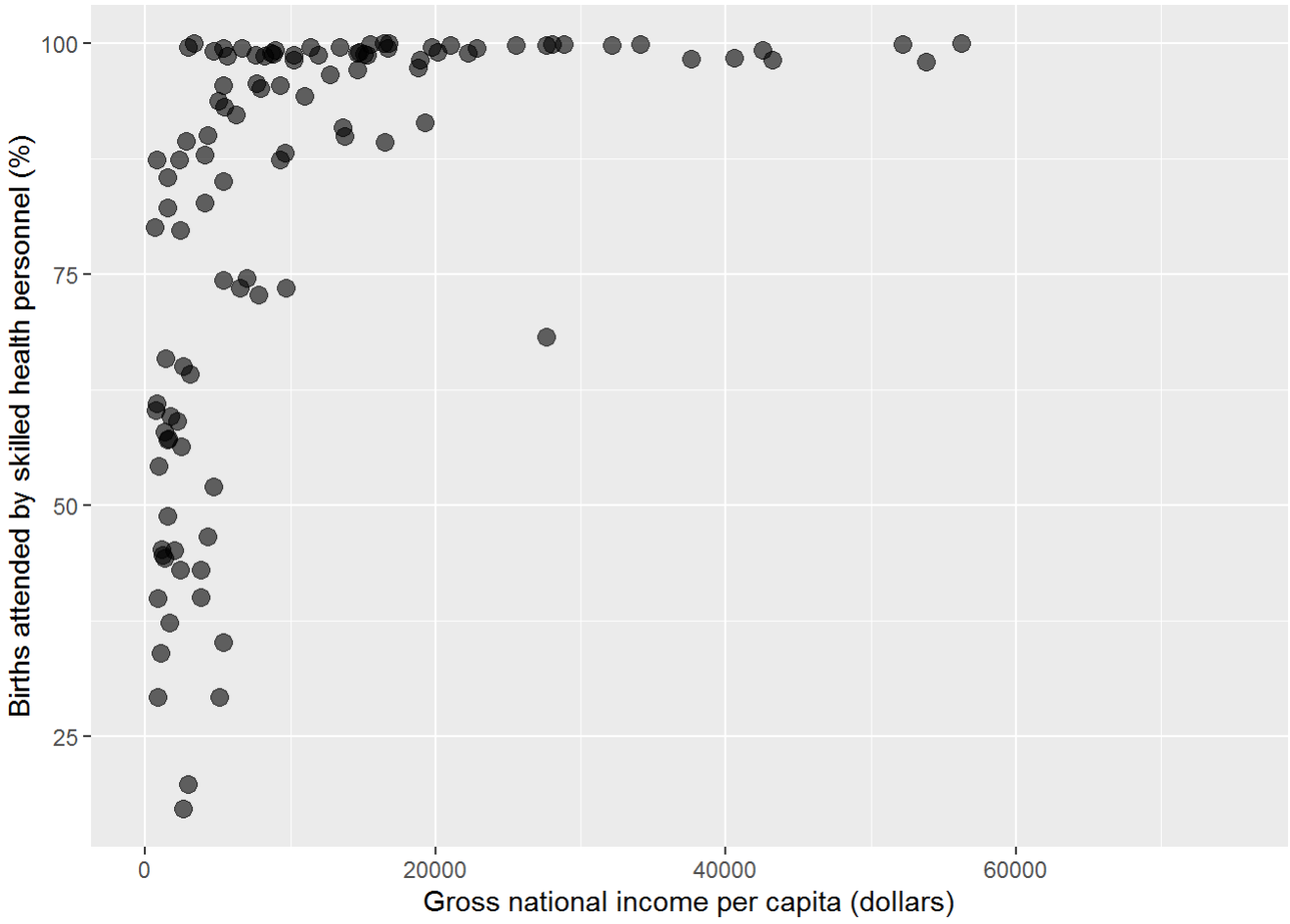
```
mortality_data <- merge_data %>% filter(Mortality.ratio != "NA")

ggplot(data = mortality_data,
       aes(x = INCOME, y = Mortality.ratio)) +
  geom_point(alpha = 0.6, size = 3) +
  xlab("Gross national income per capita (dollars)") +
  ylab("Maternal mortality ratio (per 100 000 live births)") + ylim(c(0, 1500)) + xlim(c(0, 5000
0))
```




```
birthattended_data <- merge_data %>% filter(Birth.attended != "NA")


ggplot(data = birthattended_data,
       aes(x = INCOME, y = Birth.attended)) +
  geom_point(alpha = 0.6, size = 3) +
  xlab("Gross national income per capita (dollars)") +
  ylab("Births attended by skilled health personnel (%)") + xlim(c(0, 75000))
```



TASK 5

Use the country_choropleth() function in the choroplethr package to make a world map of maternal mortality, using the most recent year for each country in the merged dataset you created. The defaults of the function will be fine; no need to tune up the plot. You can read the help file to see how the data must be formatted; you may need to rename the countries from the dataset that you've created.

```
country_year <-mortality_data %>%
        group_by(COUNTRY) %>% arrange(desc(YEAR)) %>%
        ddply("COUNTRY", summarise, YEAR = max(YEAR))

country_mortality <- left_join(country_year, mortality_data, by=c("COUNTRY" = "COUNTRY", "YEAR"
= "YEAR")) %>%  select(COUNTRY, Mortality.ratio)

country_mortality$COUNTRY <- str_to_lower(country_mortality$COUNTRY)

colnames(country_mortality) <- c("region", "value")

country_choropleth(country_mortality,
                 title = "World Map Of Maternal Mortality",
                 legend = "Maternal Mortality Ratio", num_colors = 9)
```

World Map Of Maternal Mortality