R Notebook

AUTHOR: Greeshma Chowdary Peddineni

if(!require(EDAWR)) {  
install.packages("devtools")  
devtools::install\_github("rstudio/EDAWR")  
}  
library(tidyr)  
library(dplyr)  
library(EDAWR)  
library(ggplot2)

Using the STORMS dataset, and DPLYR, answer the following: Your answer should display directly in the knitted document and include only the answer (no extraneous info) 1: show only storm (name) and date

storms %>% mutate(date = paste(year, month, day, sep = "-")) %>% select(name, date)

## # A tibble: 19,537 × 2  
## name date   
## <chr> <chr>   
## 1 Amy 1975-6-27  
## 2 Amy 1975-6-27  
## 3 Amy 1975-6-27  
## 4 Amy 1975-6-27  
## 5 Amy 1975-6-28  
## 6 Amy 1975-6-28  
## 7 Amy 1975-6-28  
## 8 Amy 1975-6-28  
## 9 Amy 1975-6-29  
## 10 Amy 1975-6-29  
## # ℹ 19,527 more rows

2: show only storms where wind was less than 55

filter(storms, wind<55)

## # A tibble: 12,546 × 13  
## name year month day hour lat long status category wind pressure  
## <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <fct> <dbl> <int> <int>  
## 1 Amy 1975 6 27 0 27.5 -79 tropical d… NA 25 1013  
## 2 Amy 1975 6 27 6 28.5 -79 tropical d… NA 25 1013  
## 3 Amy 1975 6 27 12 29.5 -79 tropical d… NA 25 1013  
## 4 Amy 1975 6 27 18 30.5 -79 tropical d… NA 25 1013  
## 5 Amy 1975 6 28 0 31.5 -78.8 tropical d… NA 25 1012  
## 6 Amy 1975 6 28 6 32.4 -78.7 tropical d… NA 25 1012  
## 7 Amy 1975 6 28 12 33.3 -78 tropical d… NA 25 1011  
## 8 Amy 1975 6 28 18 34 -77 tropical d… NA 30 1006  
## 9 Amy 1975 6 29 0 34.4 -75.8 tropical s… NA 35 1004  
## 10 Amy 1975 6 29 6 34 -74.8 tropical s… NA 40 1002  
## # ℹ 12,536 more rows  
## # ℹ 2 more variables: tropicalstorm\_force\_diameter <int>,  
## # hurricane\_force\_diameter <int>

3: split date into individual columns for month, day, year save as a new dataset called mystorms

mystorms <- storms %>%   
 unite("date", year:day, sep = "-") %>%   
 separate(date, into = c("year", "month", "day"), sep="-")  
mystorms

## # A tibble: 19,537 × 13  
## name year month day hour lat long status category wind pressure  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <fct> <dbl> <int> <int>  
## 1 Amy 1975 6 27 0 27.5 -79 tropical d… NA 25 1013  
## 2 Amy 1975 6 27 6 28.5 -79 tropical d… NA 25 1013  
## 3 Amy 1975 6 27 12 29.5 -79 tropical d… NA 25 1013  
## 4 Amy 1975 6 27 18 30.5 -79 tropical d… NA 25 1013  
## 5 Amy 1975 6 28 0 31.5 -78.8 tropical d… NA 25 1012  
## 6 Amy 1975 6 28 6 32.4 -78.7 tropical d… NA 25 1012  
## 7 Amy 1975 6 28 12 33.3 -78 tropical d… NA 25 1011  
## 8 Amy 1975 6 28 18 34 -77 tropical d… NA 30 1006  
## 9 Amy 1975 6 29 0 34.4 -75.8 tropical s… NA 35 1004  
## 10 Amy 1975 6 29 6 34 -74.8 tropical s… NA 40 1002  
## # ℹ 19,527 more rows  
## # ℹ 2 more variables: tropicalstorm\_force\_diameter <int>,  
## # hurricane\_force\_diameter <int>

4: list storm name(s) for any storms that happened in June (from your mystorms dataset)

mystorms %>% filter(month == 6) %>% select(name)

## # A tibble: 809 × 1  
## name   
## <chr>  
## 1 Amy   
## 2 Amy   
## 3 Amy   
## 4 Amy   
## 5 Amy   
## 6 Amy   
## 7 Amy   
## 8 Amy   
## 9 Amy   
## 10 Amy   
## # ℹ 799 more rows

Now using the rates dataset, answer the following:

5: what is the rate for Ethiopia in 2007?

rates %>% filter(country == "Ethiopia" & year == 2007) %>% select(rate)

## # A tibble: 1 × 1  
## rate  
## <dbl>  
## 1 4.73

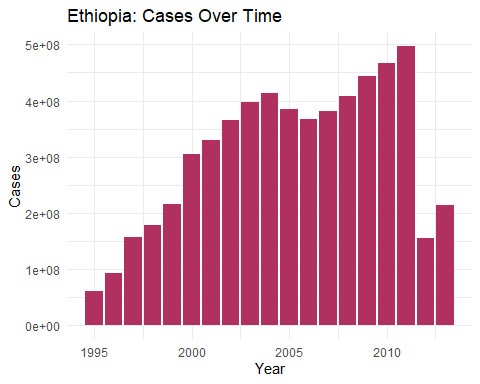
6: create a copy of the rates dataset called myrates add a new column that is population times rate, call this column ‘cases’ which country has the highest value in cases, in which year?

myrates <- rates %>% mutate(cases = population \* rate)  
highest\_cases <- myrates %>% arrange(desc(cases)) %>% slice(1)  
highest\_cases

## # A tibble: 1 × 5  
## country year cases population rate  
## <chr> <int> <dbl> <int> <dbl>  
## 1 India 2007 11985044885 1159095250 10.3

7: Change year to a factor in myrates. create a plot for Ethiopia showing cases over time

ethiopia\_cases <- myrates %>% filter(country == "Ethiopia")  
ggplot(ethiopia\_cases, aes(x = year, y = cases)) + geom\_col(fill = "maroon") + labs(title = "Ethiopia: Cases Over Time", x="Year", y = "Cases") +  
theme\_minimal()



myrates <- myrates %>% mutate(cases\_per\_1000 = cases / 1000)  
myrates

## # A tibble: 3,484 × 6  
## country year cases population rate cases\_per\_1000  
## <chr> <int> <dbl> <int> <dbl> <dbl>  
## 1 Afghanistan 1997 1331486. 19021226 0.07 1331.  
## 2 Afghanistan 1998 17742121. 19496836 0.91 17742.  
## 3 Afghanistan 1999 7395216. 19987071 0.37 7395.  
## 4 Afghanistan 2000 26568014. 20595360 1.29 26568.  
## 5 Afghanistan 2001 46324687. 21347782 2.17 46325.  
## 6 Afghanistan 2002 65054222. 22202806 2.93 65054.  
## 7 Afghanistan 2003 65187520. 23116142 2.82 65188.  
## 8 Afghanistan 2004 82384079. 24018682 3.43 82384.  
## 9 Afghanistan 2005 99443420 24860855 4 99443.  
## 10 Afghanistan 2006 124568031. 25631282 4.86 124568.  
## # ℹ 3,474 more rows