DATA 607: Tidyverse Create Assignment

James Williams

3/29/2020

Task

Your task here is to Create an Example. Using one or more TidyVerse packages, and any dataset from fivethirtyeight.com or Kaggle, create a programming sample "vignette" that demonstrates how to use one or more of the capabilities of the selected TidyVerse package with your selected dataset.

Import Data

Data comes from https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

```
library(tidyverse)
```

```
URL <- 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_
death_rates <- read_csv(URL)
head(death_rates)
```

```
## # A tibble: 6 x 72
##   `Province/State` `Country/Region`   Lat   Long `1/22/20` `1/23/20` `1/24/20`
##   <chr>            <chr>            <dbl>  <dbl>     <dbl>     <dbl>     <dbl>
## 1 <NA>             Afghanistan       33    65            0         0         0
## 2 <NA>             Albania           41.2  20.2          0         0         0
## 3 <NA>             Algeria           28.0   1.66         0         0         0
## 4 <NA>             Andorra           42.5   1.52         0         0         0
## 5 <NA>             Angola           -11.2  17.9          0         0         0
## 6 <NA>             Antigua and Bar~  17.1 -61.8          0         0         0
## # ... with 65 more variables: `1/25/20` <dbl>, `1/26/20` <dbl>,
## #   `1/27/20` <dbl>, `1/28/20` <dbl>, `1/29/20` <dbl>, `1/30/20` <dbl>,
## #   `1/31/20` <dbl>, `2/1/20` <dbl>, `2/2/20` <dbl>, `2/3/20` <dbl>,
## #   `2/4/20` <dbl>, `2/5/20` <dbl>, `2/6/20` <dbl>, `2/7/20` <dbl>,
## #   `2/8/20` <dbl>, `2/9/20` <dbl>, `2/10/20` <dbl>, `2/11/20` <dbl>,
## #   `2/12/20` <dbl>, `2/13/20` <dbl>, `2/14/20` <dbl>, `2/15/20` <dbl>,
## #   `2/16/20` <dbl>, `2/17/20` <dbl>, `2/18/20` <dbl>, `2/19/20` <dbl>,
## #   `2/20/20` <dbl>, `2/21/20` <dbl>, `2/22/20` <dbl>, `2/23/20` <dbl>,
## #   `2/24/20` <dbl>, `2/25/20` <dbl>, `2/26/20` <dbl>, `2/27/20` <dbl>,
## #   `2/28/20` <dbl>, `2/29/20` <dbl>, `3/1/20` <dbl>, `3/2/20` <dbl>,
## #   `3/3/20` <dbl>, `3/4/20` <dbl>, `3/5/20` <dbl>, `3/6/20` <dbl>,
## #   `3/7/20` <dbl>, `3/8/20` <dbl>, `3/9/20` <dbl>, `3/10/20` <dbl>,
## #   `3/11/20` <dbl>, `3/12/20` <dbl>, `3/13/20` <dbl>, `3/14/20` <dbl>,
```

```
## #   `3/15/20` <dbl>, `3/16/20` <dbl>, `3/17/20` <dbl>, `3/18/20` <dbl>,
## #   `3/19/20` <dbl>, `3/20/20` <dbl>, `3/21/20` <dbl>, `3/22/20` <dbl>,
## #   `3/23/20` <dbl>, `3/24/20` <dbl>, `3/25/20` <dbl>, `3/26/20` <dbl>,
## #   `3/27/20` <dbl>, `3/28/20` <dbl>, `3/29/20` <dbl>
```

Condition Data

```
death_sums <-death_rates %>%
  replace(is.na(.), 0) %>%
  mutate(sum = rowSums(.[5:71]))
head(death_sums)
```

```
## # A tibble: 6 x 73
##   `Province/State` `Country/Region`   Lat   Long `1/22/20` `1/23/20` `1/24/20`
##   <chr>            <chr>            <dbl>  <dbl>     <dbl>     <dbl>     <dbl>
## 1 0                Afghanistan       33    65            0         0         0
## 2 0                Albania           41.2  20.2          0         0         0
## 3 0                Algeria           28.0   1.66         0         0         0
## 4 0                Andorra           42.5   1.52         0         0         0
## 5 0                Angola           -11.2  17.9          0         0         0
## 6 0                Antigua and Bar~  17.1 -61.8          0         0         0
## # ... with 66 more variables: `1/25/20` <dbl>, `1/26/20` <dbl>,
## #   `1/27/20` <dbl>, `1/28/20` <dbl>, `1/29/20` <dbl>, `1/30/20` <dbl>,
## #   `1/31/20` <dbl>, `2/1/20` <dbl>, `2/2/20` <dbl>, `2/3/20` <dbl>,
## #   `2/4/20` <dbl>, `2/5/20` <dbl>, `2/6/20` <dbl>, `2/7/20` <dbl>,
## #   `2/8/20` <dbl>, `2/9/20` <dbl>, `2/10/20` <dbl>, `2/11/20` <dbl>,
## #   `2/12/20` <dbl>, `2/13/20` <dbl>, `2/14/20` <dbl>, `2/15/20` <dbl>,
## #   `2/16/20` <dbl>, `2/17/20` <dbl>, `2/18/20` <dbl>, `2/19/20` <dbl>,
## #   `2/20/20` <dbl>, `2/21/20` <dbl>, `2/22/20` <dbl>, `2/23/20` <dbl>,
## #   `2/24/20` <dbl>, `2/25/20` <dbl>, `2/26/20` <dbl>, `2/27/20` <dbl>,
## #   `2/28/20` <dbl>, `2/29/20` <dbl>, `3/1/20` <dbl>, `3/2/20` <dbl>,
## #   `3/3/20` <dbl>, `3/4/20` <dbl>, `3/5/20` <dbl>, `3/6/20` <dbl>,
## #   `3/7/20` <dbl>, `3/8/20` <dbl>, `3/9/20` <dbl>, `3/10/20` <dbl>,
## #   `3/11/20` <dbl>, `3/12/20` <dbl>, `3/13/20` <dbl>, `3/14/20` <dbl>,
## #   `3/15/20` <dbl>, `3/16/20` <dbl>, `3/17/20` <dbl>, `3/18/20` <dbl>,
## #   `3/19/20` <dbl>, `3/20/20` <dbl>, `3/21/20` <dbl>, `3/22/20` <dbl>,
## #   `3/23/20` <dbl>, `3/24/20` <dbl>, `3/25/20` <dbl>, `3/26/20` <dbl>,
## #   `3/27/20` <dbl>, `3/28/20` <dbl>, `3/29/20` <dbl>, sum <dbl>
```

```
data<- death_sums[, c("Country/Region", "sum")]
data<- rename(data, Country = "Country/Region", Death_Count = "sum")
head(data)
```

```
## # A tibble: 6 x 2
##   Country             Death_Count
##   <chr>                     <dbl>
## 1 Afghanistan                  17
## 2 Albania                      55
## 3 Algeria                     214
## 4 Andorra                      13
```

```
## 5 Angola                          0
## 6 Antigua and Barbuda             0
```

```
final_data <- data %>%
  group_by(Country) %>%
  summarize(sum(Death_Count))
head(final_data)
```

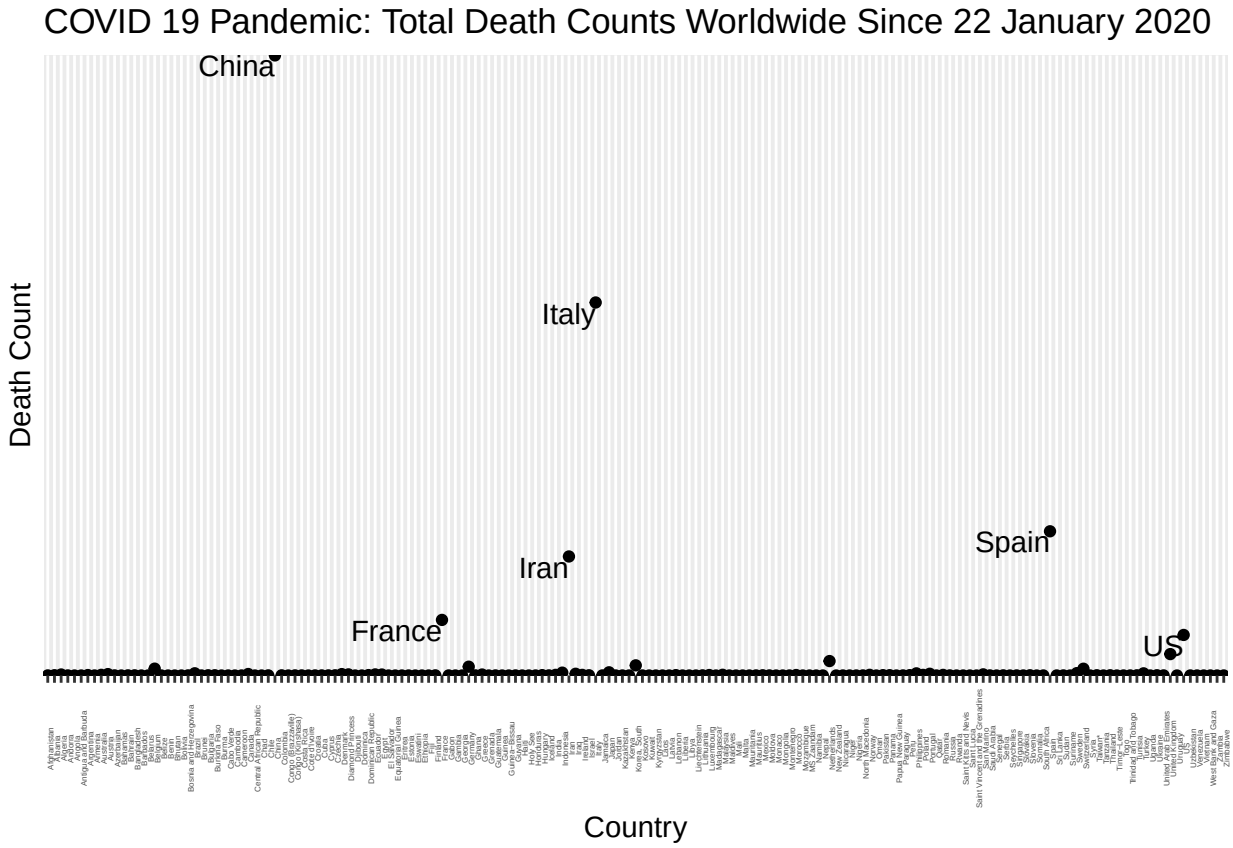
```
## # A tibble: 6 x 2
##   Country             `sum(Death_Count)`
##   <chr>                            <dbl>
## 1 Afghanistan                         17
## 2 Albania                             55
## 3 Algeria                            214
## 4 Andorra                             13
## 5 Angola                               0
## 6 Antigua and Barbuda                  0
```

```
final_data <- rename(final_data, Death_Count = `sum(Death_Count)`)
head(final_data)
```

```
## # A tibble: 6 x 2
##   Country             Death_Count
##   <chr>                     <dbl>
## 1 Afghanistan                  17
## 2 Albania                      55
## 3 Algeria                     214
## 4 Andorra                      13
## 5 Angola                        0
## 6 Antigua and Barbuda           0
```

Visualize Data

```
graph <- ggplot(final_data, aes(x = Country, y = Death_Count)) +
  geom_point() +
  theme(axis.text.x = element_text(size = 3, angle = 90))  +
  geom_text(aes(label=ifelse(Death_Count>5000, as.character(Country), '')), hjust=1, vjust=1) +
  labs(x = "Country", y = "Death Count", title = "COVID 19 Pandemic: Total Death Counts Worldwide Since
  scale_y_discrete()
graph
```



Conclusion Tidyverse features helpful tools to transform messy datasets into more accessible formats.

Resources

https://www.tidyverse.org/packages/