Graphing with ggplot2 using Wiki Ebola Data

Dr. James Wolf
School of Information Technology
Illinois State University
jrwolf@ilstu.edu

October 14, 2014

1 Reading in the data

```
> # We will be using ggplot2
> library(ggplot2)
> Country <-c("Guinea","Liberia", "Sierra Leone", "Nigeria", "Senegal")
> Cases <-c(965,3022,1753,21,1)
> Deaths <-c(623,1578,537,8,0)
> ebolaData <-data.frame(Country, Cases, Deaths)
>

> ebolaData

       Country Cases Deaths
1       Guinea   965    623
2      Liberia  3022   1578
3 Sierra Leone  1753    537
4      Nigeria    21      8
5      Senegal     1      0
```

ggplot(data=ebolaData[1:3 Above, we see a the Ebola case data by country. As you can see, two countries have many fewer cases than the other three. So, we will drop these two countries from our graph.

```
> # only using the first three countries
> ggplot(data=ebolaData[1:3,], aes(x=Country, y=Cases, fill=Country)) +
+   geom_bar(stat="identity")
>
```

Next, we want to drop the graph's guide.

```
> # no guide
> ggplot(data=ebolaData[1:3,], aes(x=Country, y=Cases, fill=Country)) +
```

```
+   geom_bar(stat="identity") +
+   guides(fill=FALSE)
>
```

Now, we want to add a title to the graph and custom legends.

```
> ggplot(data=ebolaData[1:3,], aes(x=Country, y=Cases, fill=Country)) +
+   geom_bar(stat="identity") +
+   xlab("Countries") + ylab("Reported Cases") +
+   ggtitle("Reported Ebola Cases by Country") +
+   guides(fill=FALSE)
```

Finally, we are going to create the different in a new way – which will allow us to create different graphs.

```
> # Creating the ebolaData data frame using levels
> ebolaData.new <-
+   data.frame(country = factor(c("Guinea","Liberia", "Sierra Leone", "Nigeria", "Senegal",
+                                 "Guinea","Liberia", "Sierra Leone", "Nigeria", "Senega
+                          levels=c("Liberia", "Sierra Leone", "Guinea", "Nigeria", "Senegal'
+                          stat      = factor(c("Cases","Cases","Cases","Cases","Cases",
+                                               "Deaths","Deaths","Deaths","Deaths","Deaths"),
+                                             levels=c("Cases", "Deaths")),
+                          reported  = c(965,3022,1753,21,1,623,1578,537,8,0))
>
>
```

Let's look at our new data.

```
> ebolaData.new
```

```
        country   stat reported
1        Guinea  Cases      965
2       Liberia  Cases     3022
3  Sierra Leone  Cases     1753
4       Nigeria  Cases       21
5       Senegal  Cases        1
6        Guinea Deaths      623
7       Liberia Deaths     1578
8  Sierra Leone Deaths      537
9       Nigeria Deaths        8
10      Senegal Deaths        0
```

And our new graphs – showing ebola cases and dealths side-by-side.

```
> #with the default color scheme
> ggplot(data=ebolaData.new, aes(x=country, y=reported, fill=stat)) +
+ geom_bar(stat="identity", position=position_dodge())
```

```
> #with a new color scheme
> ggplot(data=ebolaData.new, aes(x=country, y=reported, fill=stat)) +
+ geom_bar(stat="identity", position=position_dodge())    +
+ scale_fill_manual(values=c("#999999", "#E69F00"))
```