

BSA - Student Faculty Seminar

Kristin Baltrusaitis

ggplot2

ggplot2:

- based on Leland Wilkinson's Grammar of Graphics
- formal structured perspective on how to describe graphics
- references:

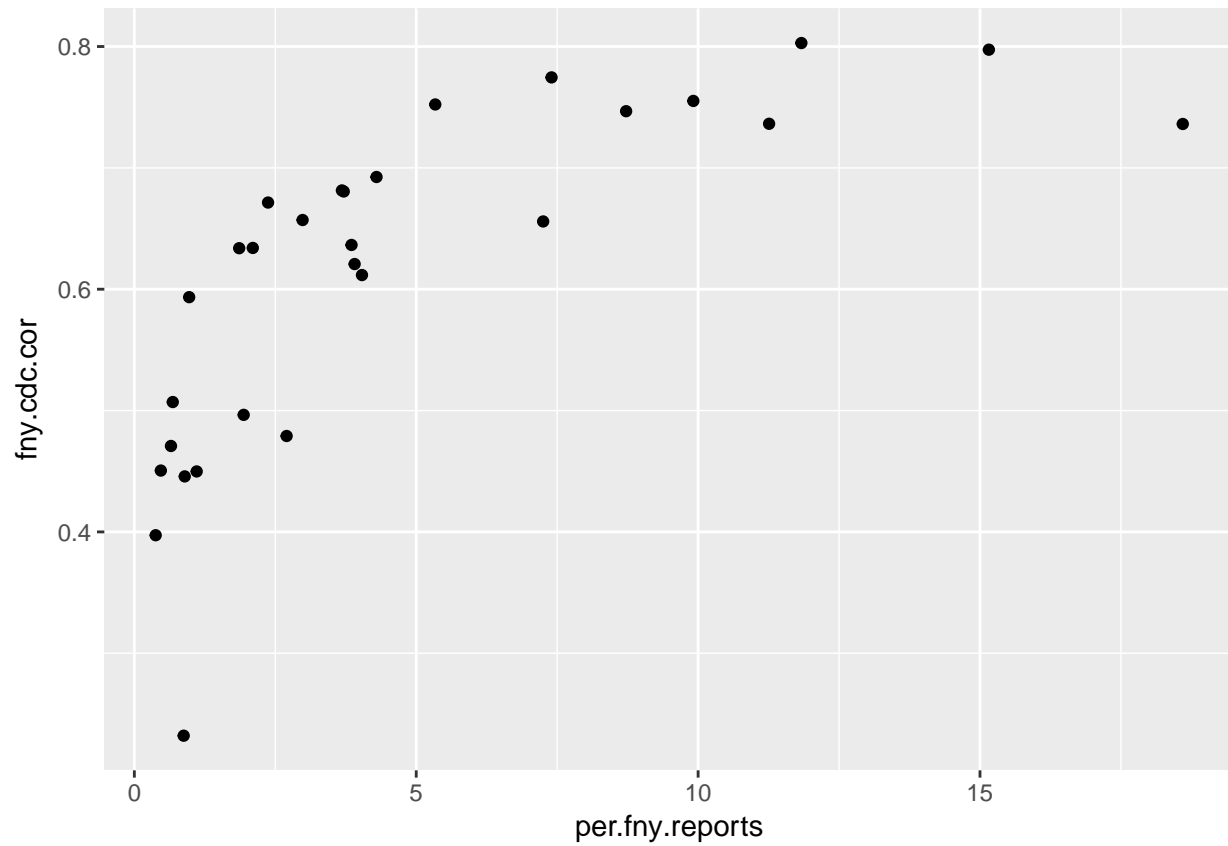
1) R Graphics Cookbook by Winston Chang

2) ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham

Example 1: Scatterplot

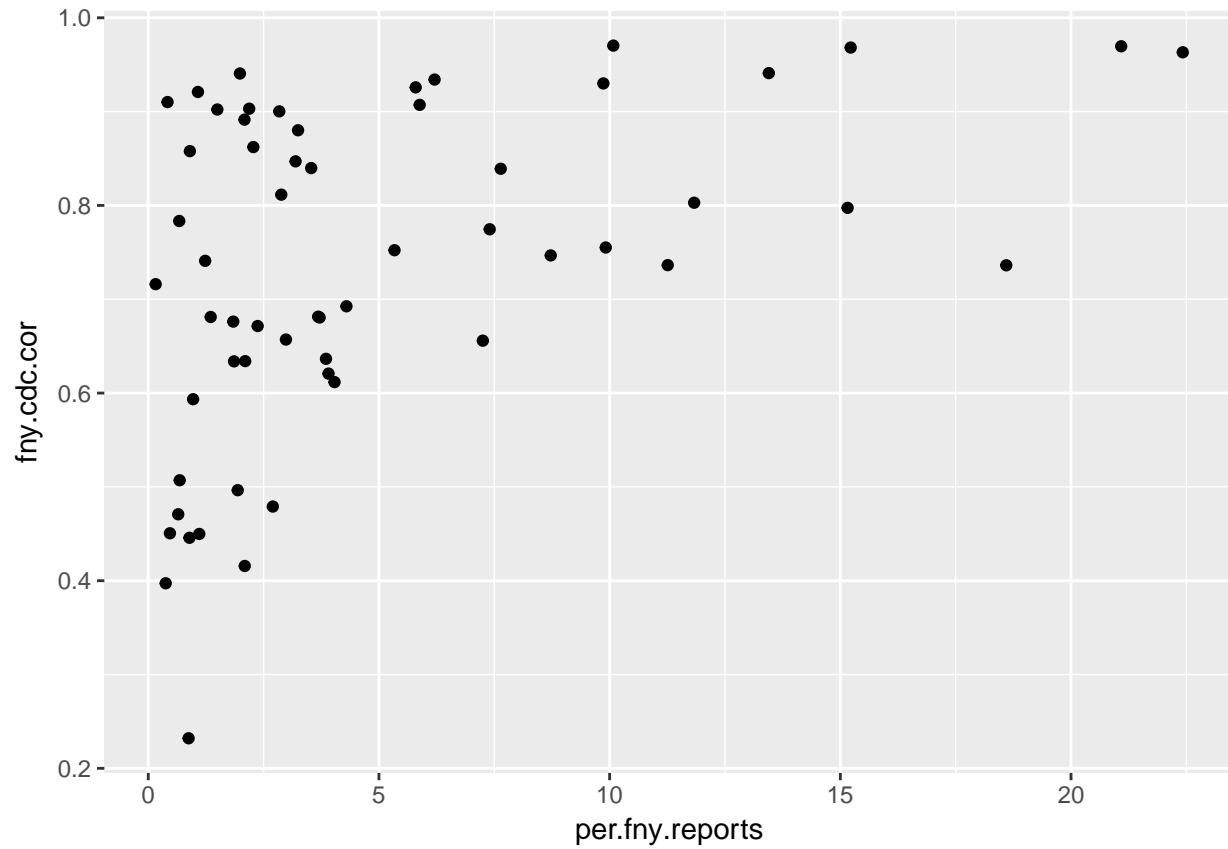
Step 1: create scatterplot

```
p1<- ggplot(data1, aes(x=per.fny.reports, y=fny.cdc.cor))+  
  geom_point()  
p1
```



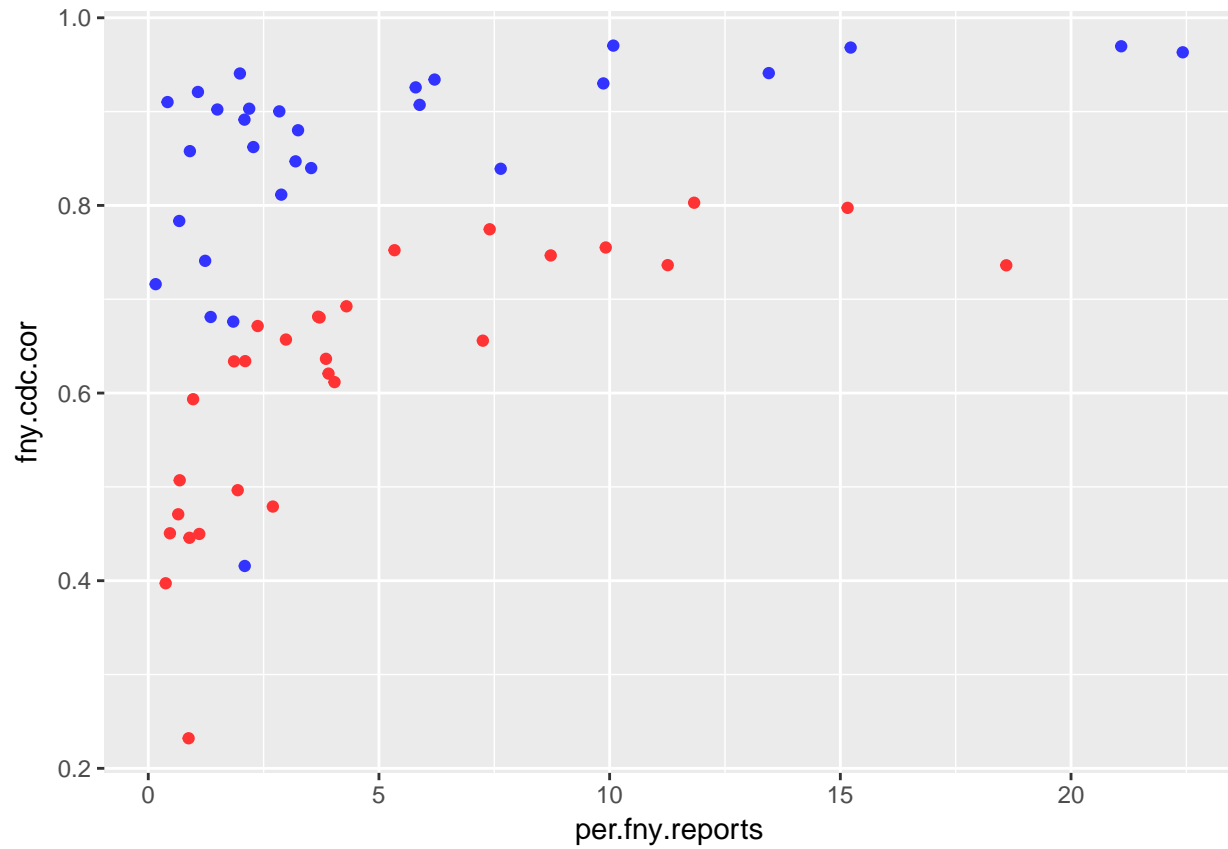
Step 2: add 2nd variable

```
p2<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor))+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor))  
p2
```



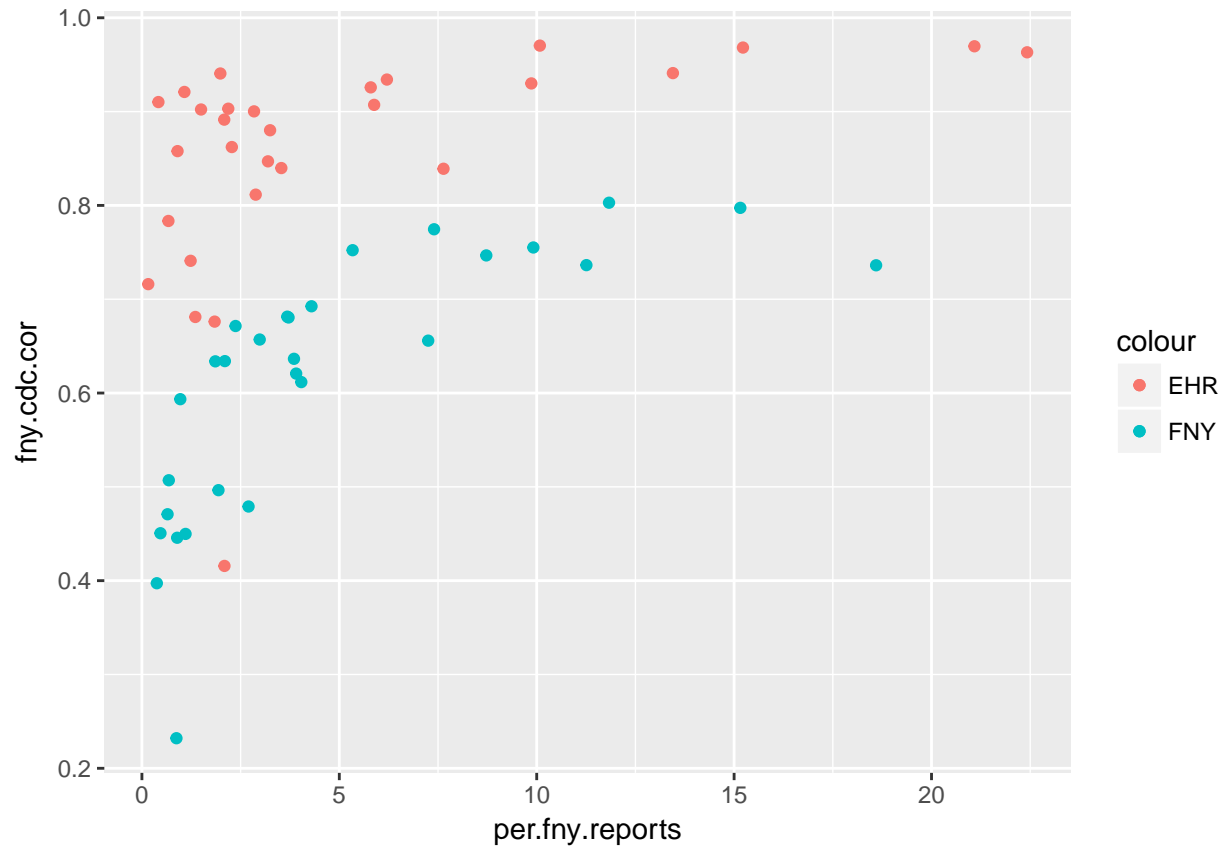
Step 3: distinguish colors chart of colors

```
p3<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor), color="#FF3333")+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor), color="#3333FF")  
p3
```



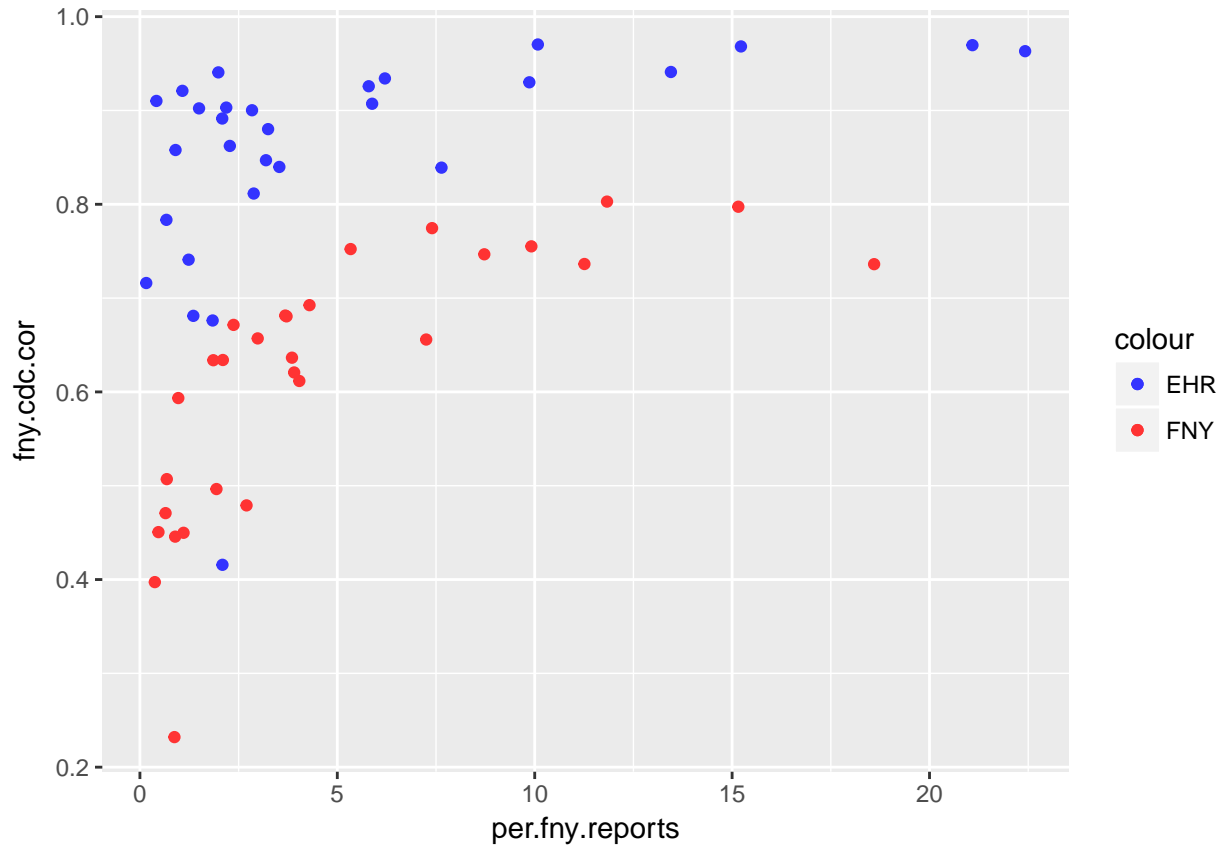
Step 4: add legend

```
p4<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY"))+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR"))  
p4
```



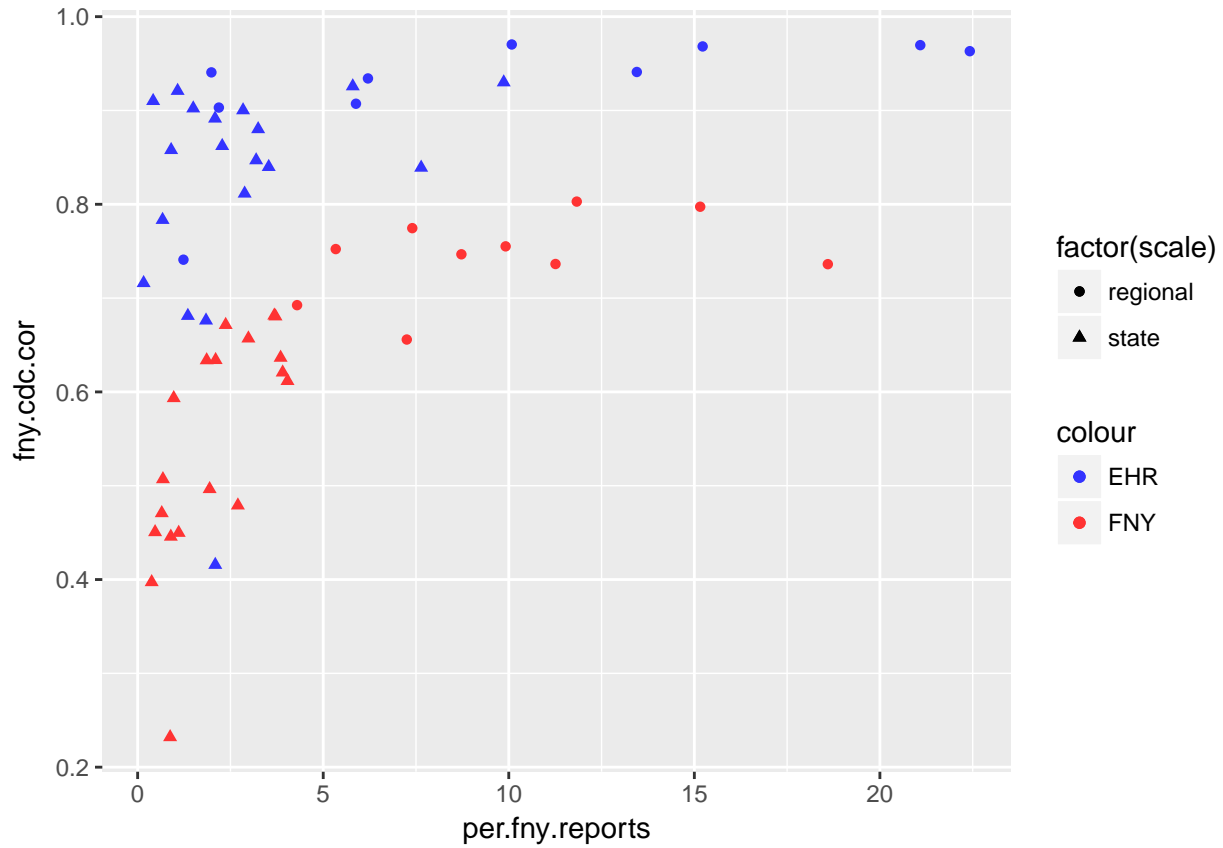
Step 5: add legend and specify colors

```
p5<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY"))+  
  geom_point(aes(x=per.athena.reports,y=athena.cdc.cor, color="EHR"))+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))  
p5
```



Step 6: add shapes to distinguish geographical resolutions

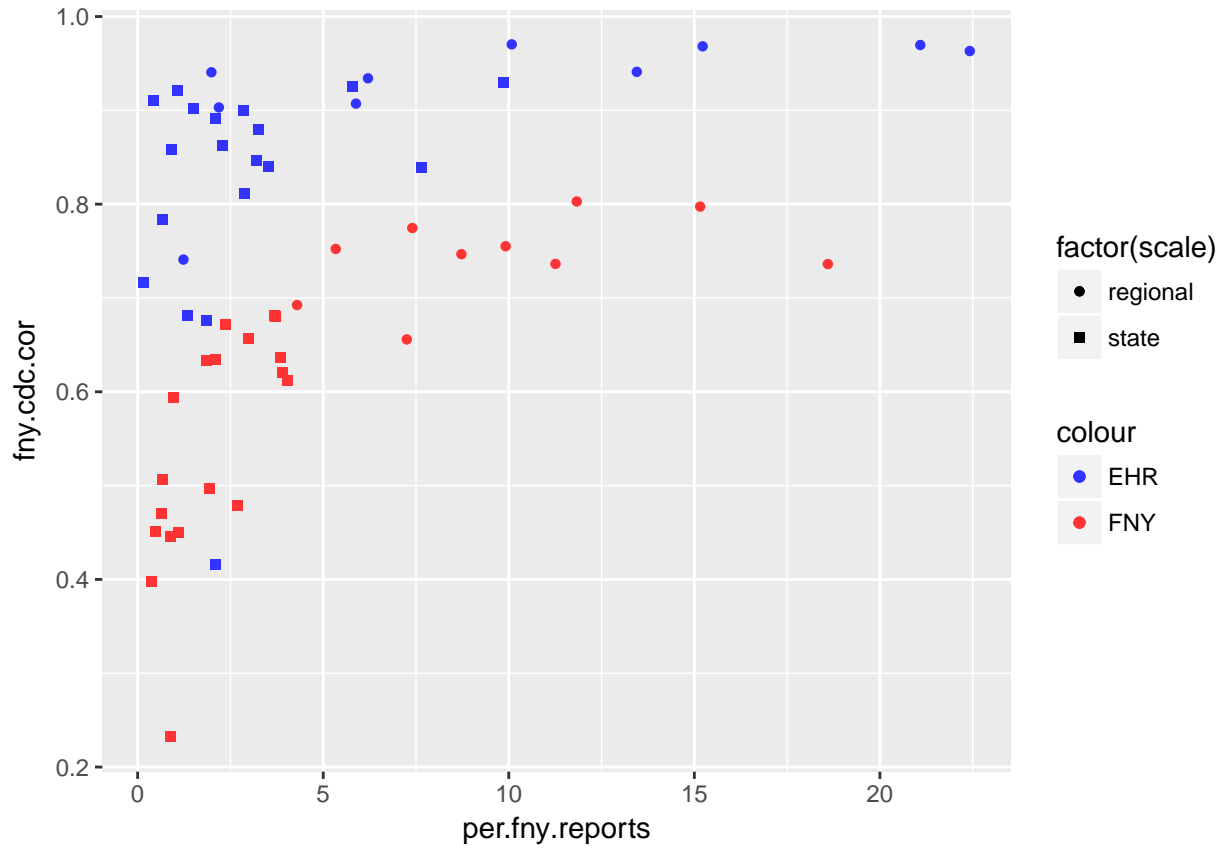
```
p6<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)))+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)))+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))  
p6
```



Step 7: change shapes list of shape options

```
p7<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)))+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)))+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+  
  scale_shape_manual(values = c(16, 15))
```

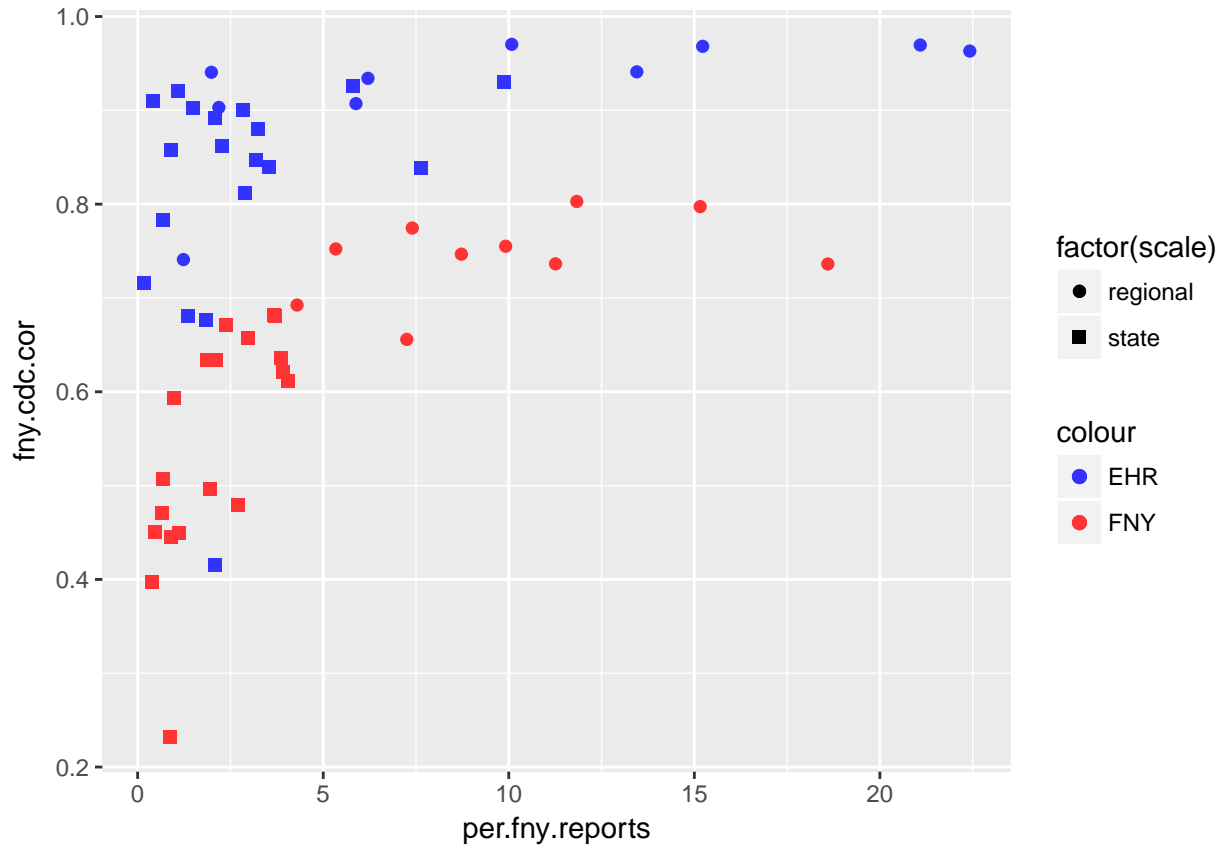
p7



Step 8: change size

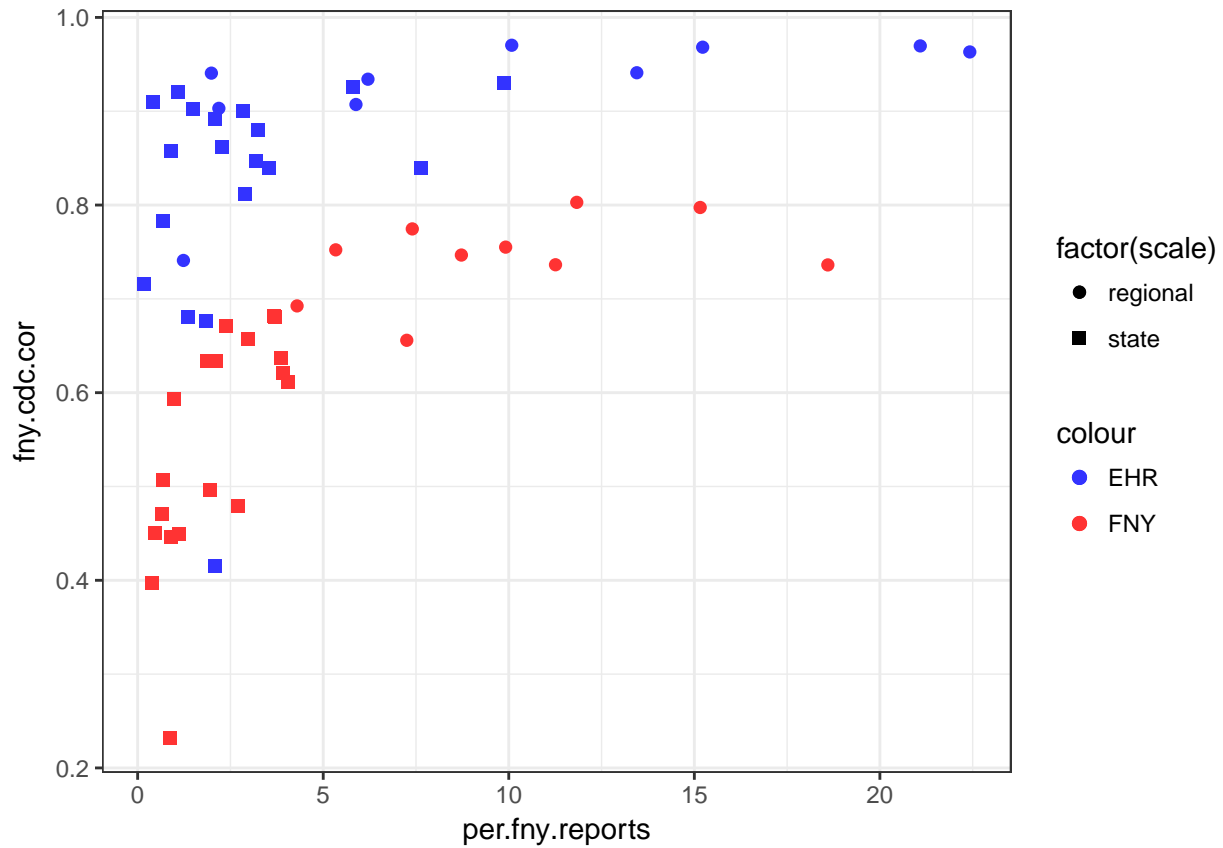
```
p8<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+  
  scale_shape_manual(values = c(16, 15))
```

p8



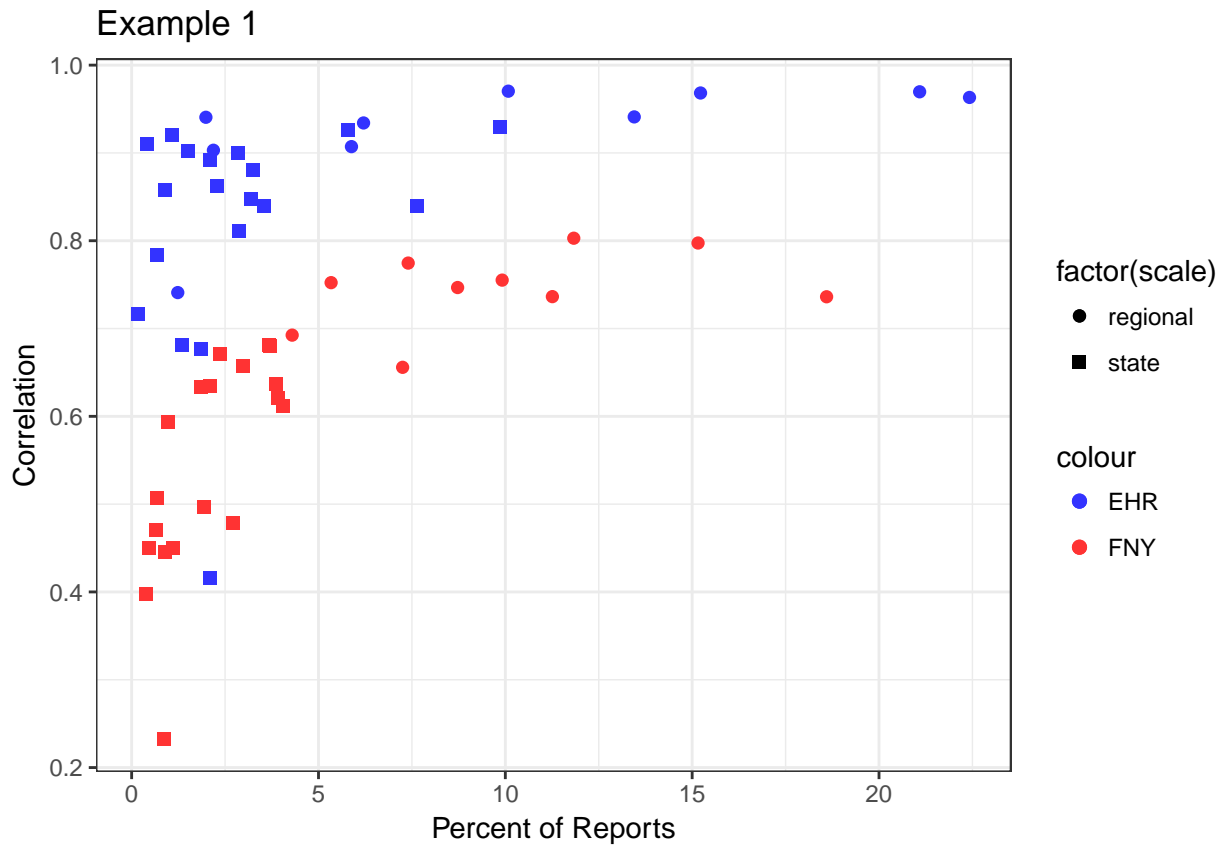
Step 9: remove grey background

```
p9<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+  
  theme_bw()+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+  
  scale_shape_manual(values = c(16, 15))  
p9
```



Step 10: add axis labels

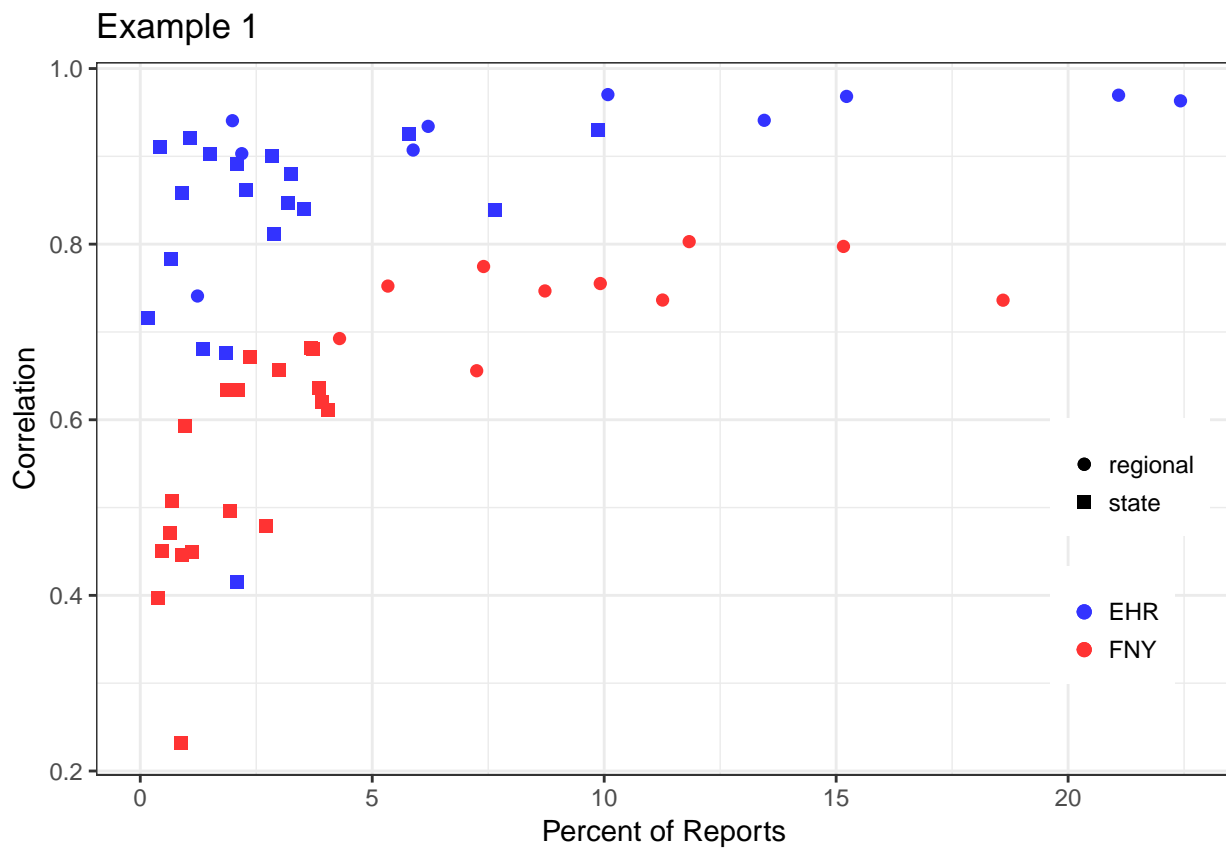
```
p10<- ggplot(data1, aes())+  
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+  
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+  
  theme_bw()+  
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+  
  scale_shape_manual(values = c(16, 15))+  
  ylab("Correlation")+xlab("Percent of Reports")+  
  ggtitle("Example 1")  
p10
```



Step 11: adjust legend

```
p11<- ggplot(data1, aes())+
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+
  theme_bw()+
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+
  scale_shape_manual(values = c(16, 15))+
  ylab("Correlation")+xlab("Percent of Reports")+
  ggtitle("Example 1") +
  theme(legend.title=element_blank(),
        legend.justification=c(1,1),
        legend.position=c(0.98,0.5),
        legend.key = element_rect(colour = "white"),
        legend.key.size = unit(0.5, "cm"))
```

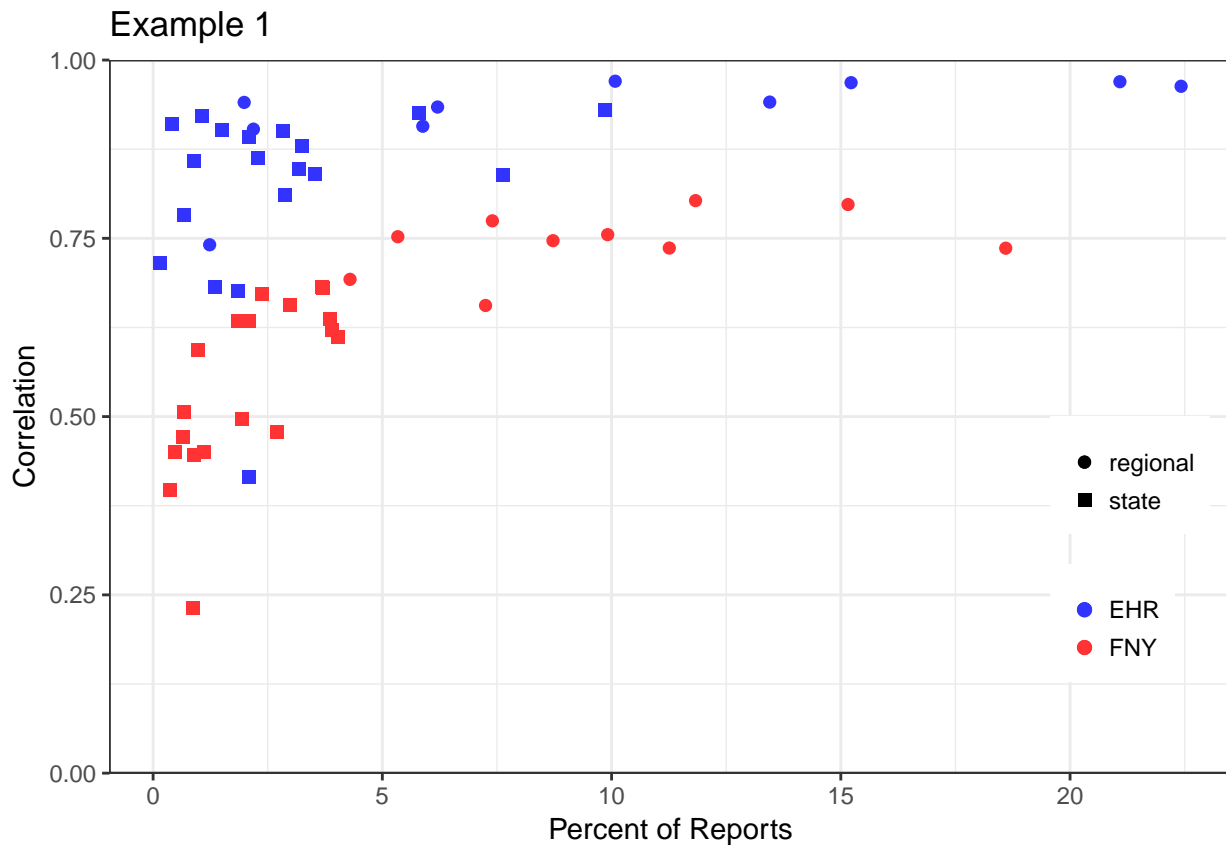
p11



Step 12: adjust y-axis

```
p12<- ggplot(data1, aes())+
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+
  theme_bw()+
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+
  scale_shape_manual(values = c(16, 15))+
  scale_y_continuous( limits = c(0,1), expand = c(0,0) )+
  ylab("Correlation")+xlab("Percent of Reports")+
  ggtitle("Example 1") +
  theme(legend.title=element_blank(),
        legend.justification=c(1,1),
        legend.position=c(0.98,0.5),
        legend.key = element_rect(colour = "white"),
        legend.key.size = unit(0.5, "cm"))
```

p12

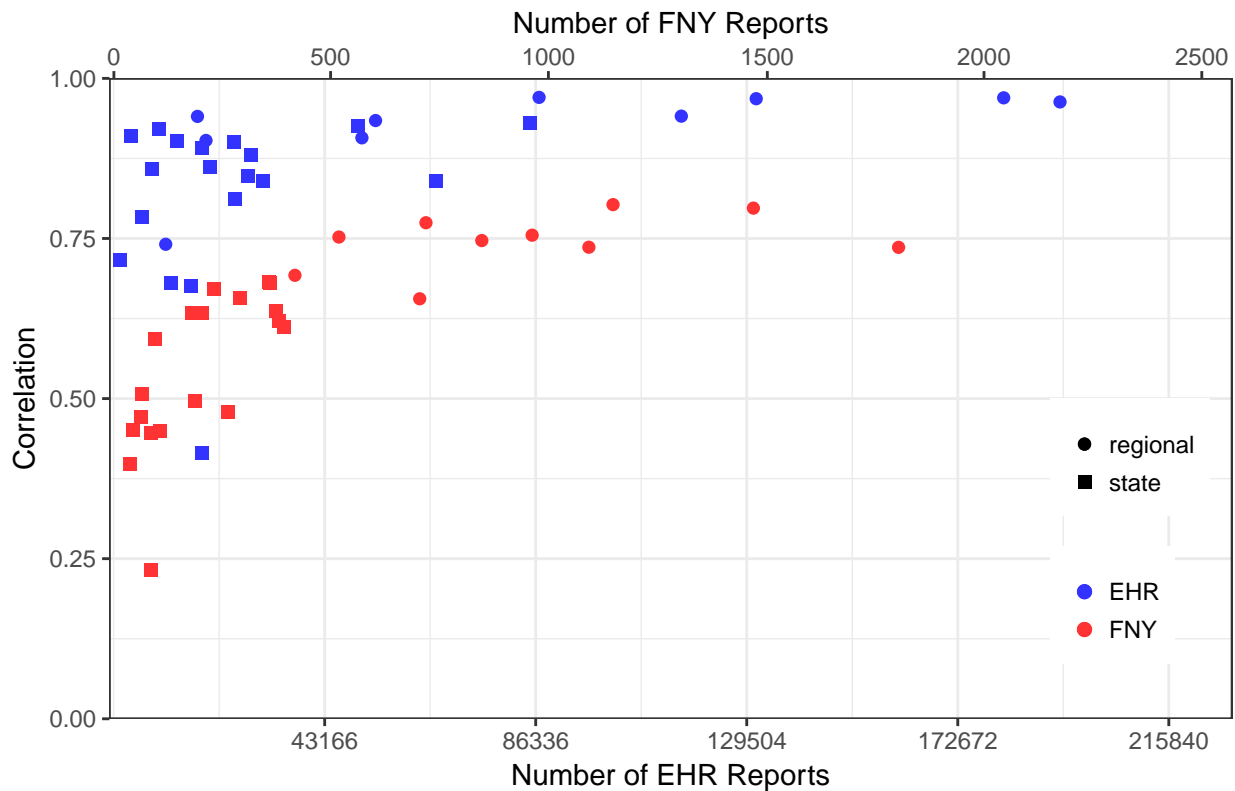


Step 13: adjust x-axis

```
p13<- ggplot(data1, aes())+
  geom_point(aes(x=per.fny.reports, y=fny.cdc.cor, color="FNY", shape = factor(scale)), size=2)+
  geom_point(aes(x=per.athena.reports, y=athena.cdc.cor, color="EHR", shape = factor(scale)), size=2)+
  theme_bw()+
  scale_color_manual(values=c(FNY="#FF3333", EHR="#3333FF"))+
  scale_shape_manual(values = c(16, 15))+
  scale_y_continuous( limits = c(0,1), expand = c(0,0) )+
  ylab("Correlation")+
  xlab("Number of EHR Reports")+
  scale_x_continuous( limits = c(-0.1,26.5), expand = c(0,0), position="bottom", breaks = c(5, 10, 15, 20),
    labels = c(43166, 86336, 129504, 172672, 215840), sec.axis= sec_axis(~.*97, name=
    "Number of FNY Reports"))
ggtitle("Example 1") +
  theme(legend.title=element_blank(),
    legend.justification=c(1,1),
    legend.position=c(0.98,0.5),
    legend.key = element_rect(colour = "white"),
    legend.key.size = unit(0.5, "cm"))
```

p13

Example 1



Save the last plot created

```
'ggsave("~/Documents/example1.pdf", width = 20, height = 20, units = "cm")'
```