Marriah\_Lewis\_HW2

Marriah Lewis

4/20/2021

#Import story-teller dataset data.storyteller <- read.csv(“~/IST707/dataset/data-storyteller.csv”, quote="") View(data.storyteller)

#Install libraries tidyverse and gridExtra library(“tidyverse”) library (“gridExtra”) #Analyzing structure and summary data str(data.storyteller) summary(data.storyteller) #Clean up data.storytellerSchool) #converting a character into a factor colnames(data.storyteller)<- c(“School”, “Section”, “VeryAhead”, “Middling”, “Behind”, “MoreBehind”, “VeryBehind”, “Completed”) #Rearrange the rows, columns, and renamed dataframe with adjusted column order col\_order<- c(“School”, “Section”, “Completed”, “VeryAhead”, “Middling”, “Behind”, “MoreBehind”, “VeryBehind”) schools<- data.storyteller[,col\_order] #No valuable data in VeryAhead schools$VeryAhead<- NULL #Focus on VeryBehind and MoreBehind schools$AbsolutelyBehind<- schoolsVeryBehind

#What is the grade distribution look like? performance<- gather(schools[,3:7]) %>% group\_by(key) %>% summarise(count=sum(value)) performance #Bar Chart barplot(performancecount, main= “Students’ Performance in Schools”, col=c(“blue”, “green”, “red”, “orange”, “yellow”, “purple”), legend= colnames(performance$count)) #Creating Frequency table; returns the frequency of each variable; which shows School A has the highest number of sections table(schools$School) #Students per school students\_per\_school<- schools %>% mutate(Num\_Students = Middling +Behind + MoreBehind + VeryBehind + Completed) %>% group\_by(School) %>% summarise(Num\_Students\_school = sum(Num\_Students)) students\_per\_school #Show visually using ggplot ggplot(students\_per\_school, aes(x=School, y=Num\_Students\_school, fill= School, label= Num\_Students\_school)) + geom\_bar(stat = “identity”) + labs(x=“Schools”, y=“Number of Students”, title= “Students by School”)

#more in depth visuals schoolsbyPerf<- gather(schools[,-c(2,8)], Status, Frequency, -School) %>% group\_by(School, Status) %>% summarise(count= sum(Frequency)) %>% mutate(pct= count/sum(count))

#Performance breakdown (stacked Plot) school\_perf\_bd<- ggplot(schoolsbyPerf, aes(x=School, y= count, fill=Status)) +geom\_bar(stat=‘identity’, position=“stack”) + labs(x=“Schools”, y=“Value”, title= “Students’ Performance Breakdown”) school\_perf\_bd #The bar chart performance breakdown shows that School A and B have high number of students middling or behind.

#Closer look at the performance via percentage school\_perf\_perc<- ggplot(schoolsbyPerf, aes(x=School, y= pct, fill=Status)) +geom\_bar(stat=‘identity’, position=“fill”) + labs(x=“Schools”, y=“Proportion”, title= “The percentage of Students’Performance”)+geom\_text(aes(label=paste(round(pct\*100, 2), ‘%’, sep=’’)), position=position\_stack(vjust=0.5), size=2)

school\_perf\_perc

#In the bar chart labeled school\_perf\_perc, School D is performing poorly by 27.27% (VeryBehind). The administration needs to take a closer look as to why students are so behind and make improvements.

#ScatterPlot- School and Section schools$SectionTot<- rowSums(schools[,3:7]) #Middling ScatterPerfMiddling<-ggplot(schools, aes(x=SectionTot, y=Middling)) + geom\_point(aes(color=School, size=Middling)) + geom\_text(aes(label=Section), size=2)+ geom\_hline(yintercept= mean(schools$Middling), linetype=“dashed”, color=“red”, size=1)+ labs(title = “Section vs Middling Counts”) ScatterPerfMiddling #In the scatterplot labeled ScatterPerfMiddling, the dotted red line indicates the mean of students in the middle category. School A Section 4, 7, and 10 and School E have the highest number of students in the middling category.

#Analyze Absolutely Behind ScatterPerfTooBehind<- ggplot(schools, aes(x=SectionTot, y=AbsolutelyBehind))+ geom\_point(aes(color=School, size=AbsolutelyBehind)) + geom\_text(aes(label=Section), size=2)+ geom\_hline(yintercept= mean(schools$AbsolutelyBehind), linetype=“dashed”, color= “orange”, size=1)+ labs(title= “Section vs AbsolutelyBehind Counts” ) ScatterPerfTooBehind #School A and E have high numbers of students that are absolutely behind but there is still rooom for improvement if the adminstration focuses on these two schools.