Untitled

Georges Colbert

2/23/2018

library(readxl)  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readxl)  
library(ggplot2)  
library(dplyr)  
  
  
################################ CCU01C MISSION: Third Grade  
  
  
################ this is the data taken from CCU01C class  
  
Summary <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = 'Summary')  
Mission.specific <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = "Mission-specific")  
Exercises <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = "Exercises")  
Videos <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = "Videos")  
Points <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = "Points")  
Badges <- read\_excel("Khan Academy Report -- 02-17-1970 to 02-28-2018.xlsx", sheet = "Badges")  
Skills <- read\_excel("Khan Academy Report Skill Progress.xlsx")  
  
  
  
####Filter for only the exercises that are part of the 3rd grade mission   
  
grade3.ex <- Exercises[Exercises$Exercise %in% Skills$Exercise,]

What is the one exercise in the 3rd grade mission that no student has yet attempted?

The number of exercises the students attempted is 114,but the total Skills in the Third Grade missions is 115.

"Telling time on the number line" is the one skill no student has yet attempted.

#what is the one exercise in the 3rd grade mission that no student has yet attempted?  
  
nrow(Skills)

## [1] 115

length(unique(grade3.ex$Exercise))

## [1] 114

n <- unique(grade3.ex$Exercise)   
p <- Skills$Exercise  
  
  
setdiff(p,n)

## [1] "Telling time on the number line"

######## Create a total minutes spent on Mission Collumn  
  
ccu01c <- Mission.specific  
  
m <- grade3.ex %>% select(Student,`Time Spent (min)`) %>% group\_by(Student) %>% summarise(total.minutes = sum(`Time Spent (min)`))  
  
m <- as.data.frame(m)  
  
ccu01c <- full\_join(ccu01c,m, by="Student")  
  
#####Creating a totalHours Collumn  
  
  
myHours <- function (x) {  
 return (x/60)  
}  
  
  
  
totalHours <- lapply(ccu01c$total.minutes, myHours)  
totalHours <- unlist(totalHours)  
ccu01c <- cbind(ccu01c,totalHours)

Summary of hours spent by students on the Third grade mission

Maximum hours a student has spent on the platform is 39.5, the average hours students have spent is 15, the median is 11. the minimum is 0.

#### Summarize of totalHours  
  
summarise(ccu01c, maxHours=max(ccu01c$totalHours,na.rm=TRUE), medianhours=median(ccu01c$totalHours,na.rm=TRUE), meanHours=mean(ccu01c$totalHours,na.rm=TRUE))

## maxHours medianhours meanHours  
## 1 39.55861 11.39806 15.40182

summary of weeks spent by students on the third grade mission

Maximum amount of weeks, as defined by 3 hours per week, is 13 weeks on the 3rd grade mission. the median amount of weeks spent by students is 4, the average is 5. the minimum is 0.

###### Creating a Weeks Performing Collumn  
  
myWeeks <- function (x) {  
 return (x/3)  
}  
  
totalWeeks <- lapply(ccu01c$totalHours, myWeeks)  
totalWeeks <- unlist(totalWeeks)  
ccu01c<- cbind(ccu01c,totalWeeks)  
  
  
  
summarise(ccu01c, max=max(ccu01c$totalWeeks,na.rm=TRUE), median=median(ccu01c$totalWeeks,na.rm=TRUE), mean=mean(ccu01c$totalWeeks,na.rm=TRUE))

## max median mean  
## 1 13.1862 3.799352 5.133941

Time spent per exercise

the Maximum amount of time a student has spent on a exercise was 368.2, and the exercise was "Read bar graphs and solve 1-step problems".

the average amount of time students have spent on a exercise is 12.27.

########### Time Spent on Exercises  
  
##### Overall Numbers for Time Spent on mission exercises  
  
summarise(grade3.ex, max = max(`Time Spent (min)`),  
 min = min(`Time Spent (min)`),  
 avg = mean(`Time Spent (min)`),)

## # A tibble: 1 x 3  
## max min avg  
## <dbl> <dbl> <dbl>  
## 1 368.2833 0 12.27166

### Time spent by  
i <- grade3.ex %>% select(Student,`Time Spent (min)` ) %>% group\_by(Student) %>% summarise(Ex.time.max = max(`Time Spent (min)`),  
 Ex.time.min = min(`Time Spent (min)`),  
 ex.time.avg = mean(`Time Spent (min)`),  
 ex.time.median = median(`Time Spent (min)`))   
  
  
  
##### Summarize time spent per Excercise  
  
 d <- grade3.ex %>% select(Exercise,`Time Spent (min)` ) %>% group\_by(Exercise) %>% summarise(Ex.time.max = max(`Time Spent (min)`),  
 Ex.time.min = min(`Time Spent (min)`),  
 ex.time.avg = mean(`Time Spent (min)`),  
 ex.time.median = median(`Time Spent (min)`))   
i <- as.data.frame(i)  
  
ccu01c <- full\_join(ccu01c,i, by="Student")  
  
  
##### Create a progress collumn  
  
nSkills <- nrow(Skills)  
  
prog <- function (a,b,c,d,e) {  
 mastered <- a \*4  
 level2 <- b \*3  
 level1 <- c \* 1  
 practiced <- d\*1  
 per <- (mastered+level2+level1+practiced)/(4\*e)  
 return(per)  
   
}  
  
  
Progress <- mapply(prog, Mission.specific$Mastered,  
 Mission.specific$`Level 2`, Mission.specific$`Level 1`, Mission.specific$Practiced, nSkills)  
  
ccu01c <- cbind(ccu01c,Progress)  
  
  
  
  
ccu01c <- ccu01c[order(ccu01c$Progress, decreasing = TRUE),]  
  
  
head(ccu01c)

## Student Struggling Needs Practice Practiced Level 1 Level 2 Mastered  
## 3 ccu01s124 0 0 0 0 0 104  
## 24 ccu01s150 0 0 0 0 0 104  
## 25 ccu01s151 0 0 0 0 0 104  
## 26 ccu01s152 2 0 0 0 0 102  
## 13 ccu01s134 13 1 0 0 2 98  
## 11 ccu01s132 11 3 0 1 1 97  
## total.minutes totalHours totalWeeks Ex.time.max Ex.time.min ex.time.avg  
## 3 1614.633 26.91056 8.970185 110.0167 0.00 15.52532  
## 24 1725.267 28.75444 9.584815 101.6500 0.15 16.58910  
## 25 1403.533 23.39222 7.797407 91.4500 0.00 13.49551  
## 26 1602.067 26.70111 8.900370 125.5500 0.20 15.40449  
## 13 2288.733 38.14556 12.715185 243.9500 0.00 20.07661  
## 11 2373.517 39.55861 13.186204 107.8333 0.00 21.00457  
## ex.time.median Progress  
## 3 5.033333 0.9043478  
## 24 9.458333 0.9043478  
## 25 5.850000 0.9043478  
## 26 7.758333 0.8869565  
## 13 11.400000 0.8652174  
## 11 13.016667 0.8521739

The following table is the top 5 exercises that have the most student with a struggling status.

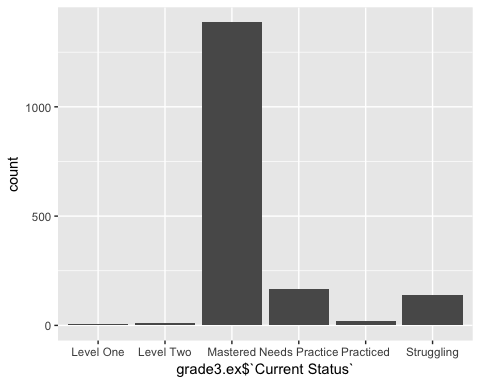
###################### Find the exercise students struggle with the most  
  
z <- filter(grade3.ex, `Current Status`=='Struggling')  
  
z <- table(z$Exercise)  
  
z<- as.data.frame(z)  
  
x <- filter(z, Freq>5)  
  
x

## Var1 Freq  
## 1 Categorize quadrilaterals 9  
## 2 Compare fractions of different wholes 6  
## 3 Find a missing side length when given perimeter 7  
## 4 Meaning of division 6  
## 5 Read bar graphs and solve 2 step problems 9  
## 6 Read picture graphs 9  
## 7 Read picture graphs (multi-step problems) 7

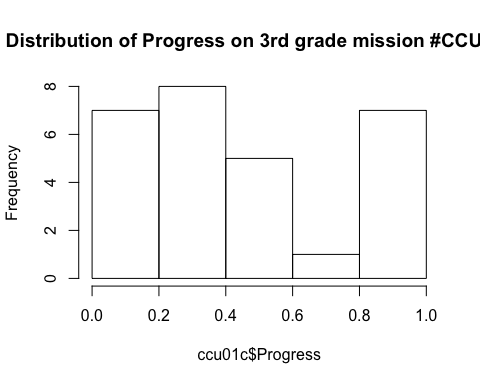
###################################################### Graphs ######################################################   
  
######### breakdown of current status  
(bean <- table(grade3.ex$`Current Status`))

##   
## Level One Level Two Mastered Needs Practice Practiced   
## 7 12 1386 168 21   
## Struggling   
## 138

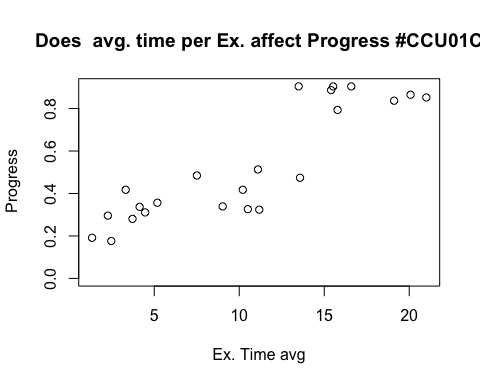
ggplot(grade3.ex,aes(grade3.ex$`Current Status`))+geom\_bar()



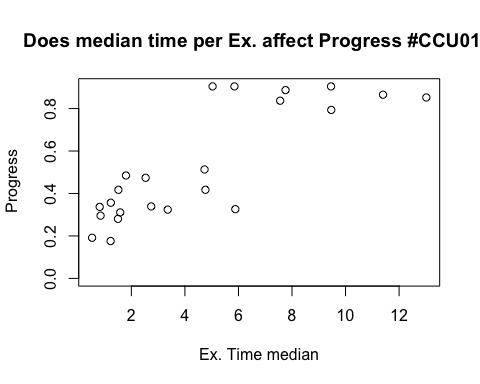
hist(ccu01c$Progress, main= " Distribution of Progress on 3rd grade mission #CCU01C")



plot(ccu01c$ex.time.avg,ccu01c$Progress, main= "Does avg. time per Ex. affect Progress #CCU01C", xlab="Ex. Time avg", ylab= " Progress" )

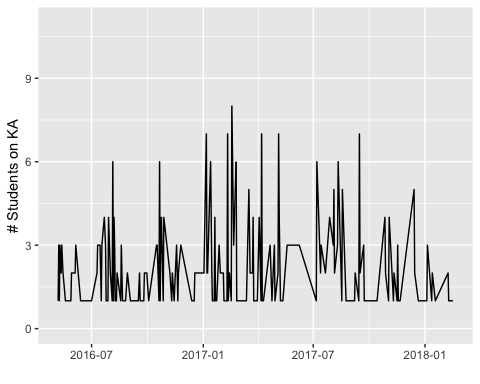


plot(ccu01c$ex.time.median,ccu01c$Progress, main= "Does median time per Ex. affect Progress #CCU01C", xlab="Ex. Time median", ylab= " Progress" )



######################## Time Series   
  
  
##### Time series of Student participation per day  
  
grade3.ex$`Last Done` <- as.Date(grade3.ex$`Last Done`)   
  
  
d <- grade3.ex %>% select(`Last Done`,Student ) %>% group\_by(`Last Done`) %>% summarise(nstudents= n\_distinct(Student))  
  
  
ggplot(d, aes(`Last Done`, nstudents)) + geom\_line() +  
 scale\_x\_date() + xlab(" ") + ylab(" # Students on KA ") + coord\_cartesian(ylim = c(0, 11))

## Warning: Removed 1 rows containing missing values (geom\_path).



###### Time series of median time spent on KA per day  
  
  
c <- grade3.ex %>% select(`Last Done`,`Time Spent (min)` ) %>% group\_by(`Last Done`) %>% summarise(avg.time= median(`Time Spent (min)`))   
  
  
  
  
ggplot(c, aes(`Last Done`, avg.time)) + geom\_line() +  
 scale\_x\_date() + xlab(" ") + ylab(" Median Time Spent ")

## Warning: Removed 1 rows containing missing values (geom\_path).

