UMCH

Cora Boyoung Jung, Enoch Mwesigwa, Jordan Severn

November 13, 2020

Import csv & Change data type

```
UMCHdata <- read_csv("https://raw.githubusercontent.com/Cora-Boyoung-Jung/UMCH/main/data/UMCH.csv",
                 col_types = cols(Birthdate = col_date(format = "%m/%d/%Y"),
                                  AgeYear = col_integer(),
                                  AgeMonth = col_integer(),
                                  AgeDay = col_integer()))

glimpse(UMCHdata)
```

```
## Rows: 33
## Columns: 17
## $ Filename                       <chr> "failed_infant_1", "failed_infant_2",...
## $ Birthdate                      <date> 2020-05-26, 2020-06-03, 2020-03-05, ...
## $ AgeGroup                       <chr> "infant", "infant", "infant", "infant...
## $ Age                            <chr> "0 Years, 4 Months, 27  Days", "0 Yea...
## $ AgeYear                        <int> 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 1...
## $ AgeMonth                       <int> 4, 4, 7, 11, 3, 9, 7, 8, 8, 5, 6, 2, ...
## $ AgeDay                         <int> 27, 20, 18, 10, 22, 27, 8, 29, 15, 3,...
## $ PhysicalDevelopment            <dbl> 4.0, 4.0, 13.0, 23.0, 6.0, 19.0, 19.0...
## $ LanguageDevelopment            <dbl> 7.0, 5.0, 11.0, 32.0, 11.0, 21.0, 15....
## $ Adaptive_SelfHelp              <dbl> 3, 3, 4, 6, 3, 8, 8, 8, 6, 7, 8, 8, 8...
## $ Adaptive_SocialEmotional       <dbl> 3, 0, 5, 14, 5, 12, 12, 12, 12, 11, 1...
## $ AcademicAndCognitive           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ AcademicAndCognitive_Maths     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ AcademicAndCognitive_Literacy <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ TotalScore                     <dbl> 17.0, 12.0, 33.0, 75.0, 25.0, 60.0, 5...
## $ Status                         <chr> "failed", "failed", "failed", "passed...
## $ Examiner                       <chr> "Sam McGowen", "Sam McGowen", "Sam Mc...
```

Tidying data

```
UMCH <- UMCHdata %>% mutate(Status = tolower(Status))
neworder <- c("infant","toddler","two_year", "three_year", "four_year")
library(plyr)  ## or dplyr (transform -> mutate)
UMCH <- arrange(transform(UMCH,
             AgeGroup=factor(AgeGroup,levels=neworder)),AgeGroup)
```

Exploring data set

```
head(UMCH)
```

```
##           Filename  Birthdate AgeGroup                           Age AgeYear
## 1  failed_infant_1 2020-05-26   infant  0 Years, 4 Months, 27  Days       0
## 2  failed_infant_2 2020-06-03   infant  0 Years, 4 Months, 20  Days       0
## 3 failed_infant_3* 2020-03-05   infant  0 Years, 7 Months, 18  Days       0
## 4  passed_infant_1 2019-11-13   infant 0 Years, 11 Months, 10  Days       0
## 5  passed_infant_2 2020-07-01   infant  0 Years, 3 Months, 22  Days       0
## 6 failed_toddler_1 2018-12-26  toddler  1 Years, 9 Months, 27  Days       1
##   AgeMonth AgeDay PhysicalDevelopment LanguageDevelopment Adaptive_SelfHelp
## 1        4     27                  4                   7                 3
## 2        4     20                  4                   5                 3
## 3        7     18                 13                  11                 4
## 4       11     10                 23                  32                 6
## 5        3     22                  6                  11                 3
## 6        9     27                 19                  21                 8
##   Adaptive_SocialEmotional AcademicAndCognitive AcademicAndCognitive_Maths
## 1                        3                   NA                         NA
## 2                        0                   NA                         NA
## 3                        5                   NA                         NA
## 4                       14                   NA                         NA
## 5                        5                   NA                         NA
## 6                       12                   NA                         NA
##   AcademicAndCognitive_Literacy TotalScore Status        Examiner
## 1                            NA         17 failed     Sam McGowen
## 2                            NA         12 failed     Sam McGowen
## 3                            NA         33 failed     Sam McGowen
## 4                            NA         75 passed Melissa Swanson
## 5                            NA         25 passed     Sam McGowen
## 6                            NA         60 failed      Ms. Jentle
```

```
summary(UMCH)
```

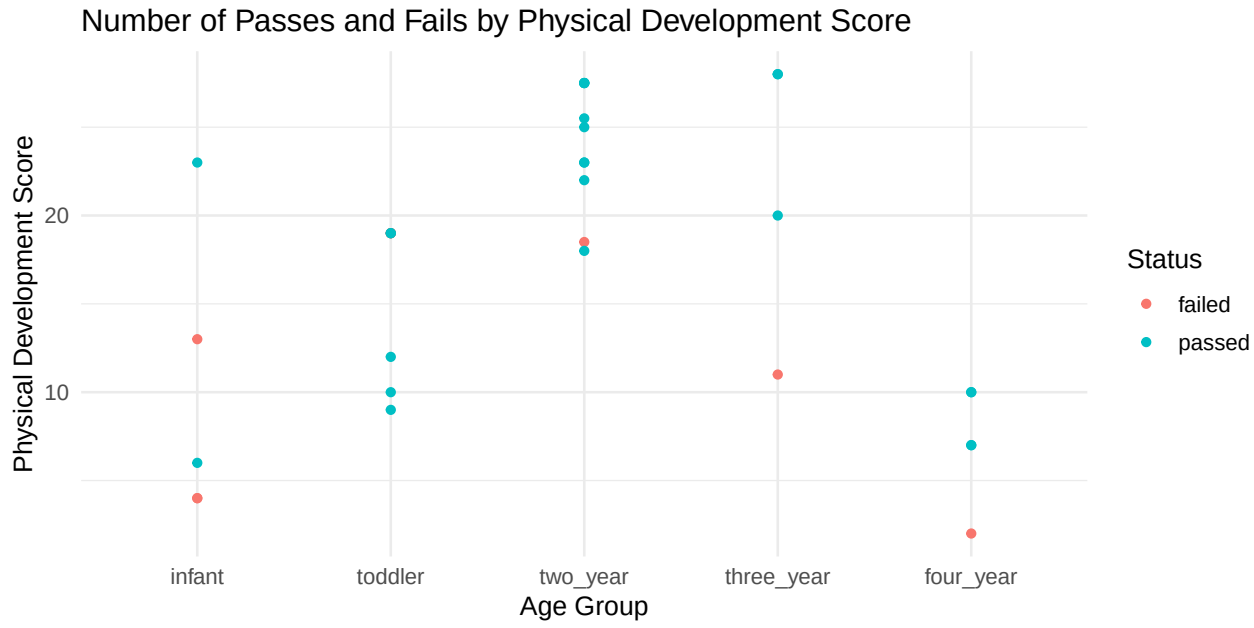
```
##    Filename           Birthdate                AgeGroup      Age
##  Length:33          Min.   :2016-02-16   infant    : 5   Length:33
##  Class :character   1st Qu.:2017-07-09   toddler   : 8   Class :character
##  Mode  :character   Median :2018-04-10   two_year  :11   Mode  :character
##                     Mean   :2018-04-21   three_year: 4
##                     3rd Qu.:2019-02-08   four_year : 5
##                     Max.   :2020-07-01
##
##     AgeYear         AgeMonth          AgeDay      PhysicalDevelopment
##  Min.   :0.000   Min.   : 1.000   Min.   : 1.00   Min.   : 2.00
##  1st Qu.:1.000   1st Qu.: 4.000   1st Qu.:11.00   1st Qu.:10.00
##  Median :2.000   Median : 7.000   Median :15.00   Median :19.00
##  Mean   :1.939   Mean   : 6.242   Mean   :16.58   Mean   :17.09
##  3rd Qu.:3.000   3rd Qu.: 8.000   3rd Qu.:25.00   3rd Qu.:23.00
##  Max.   :4.000   Max.   :11.000   Max.   :29.00   Max.   :28.00
##
##  LanguageDevelopment Adaptive_SelfHelp Adaptive_SocialEmotional
##  Min.   : 5.00       Min.   :3.000     Min.   : 0.000
##  1st Qu.:21.00       1st Qu.:4.000     1st Qu.: 5.000
```

```
##  Median :35.50       Median :7.000     Median :12.000
##  Mean   :32.09       Mean   :6.154     Mean   : 9.385
##  3rd Qu.:44.00       3rd Qu.:8.000     3rd Qu.:12.000
##  Max.   :51.00       Max.   :8.000     Max.   :14.000
##                      NA's   :20        NA's   :20
##  AcademicAndCognitive AcademicAndCognitive_Maths AcademicAndCognitive_Literacy
##  Min.   : 2.50        Min.   : 5.0               Min.   : 2.0
##  1st Qu.: 7.50        1st Qu.:15.0               1st Qu.: 9.0
##  Median :13.25        Median :15.5               Median :10.0
##  Mean   :13.22        Mean   :14.1               Mean   : 8.2
##  3rd Qu.:18.38        3rd Qu.:17.0               3rd Qu.:10.0
##  Max.   :21.50        Max.   :18.0               Max.   :10.0
##  NA's   :13           NA's   :28                 NA's   :28
##    TotalScore      Status            Examiner
##  Min.   :12.0   Length:33          Length:33
##  1st Qu.:54.0   Class :character   Class :character
##  Median :74.0   Mode  :character   Mode  :character
##  Mean   :66.7
##  3rd Qu.:86.0
##  Max.   :95.5
##
```

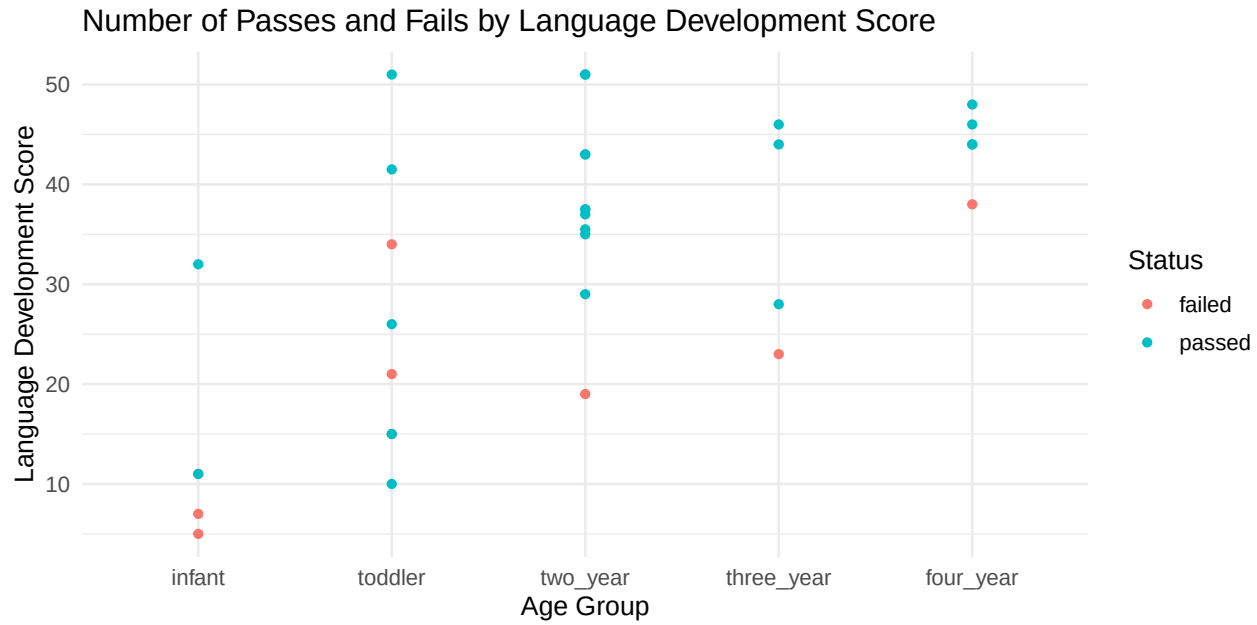
Draft Graphics

- Which domain in the areas of development is scored the lowest and highest in which age group and overall?

```
ggplot(UMCH, aes(x = AgeGroup,  y = PhysicalDevelopment, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Physical Development Score",
       title = "Number of Passes and Fails by Physical Development Score")
```

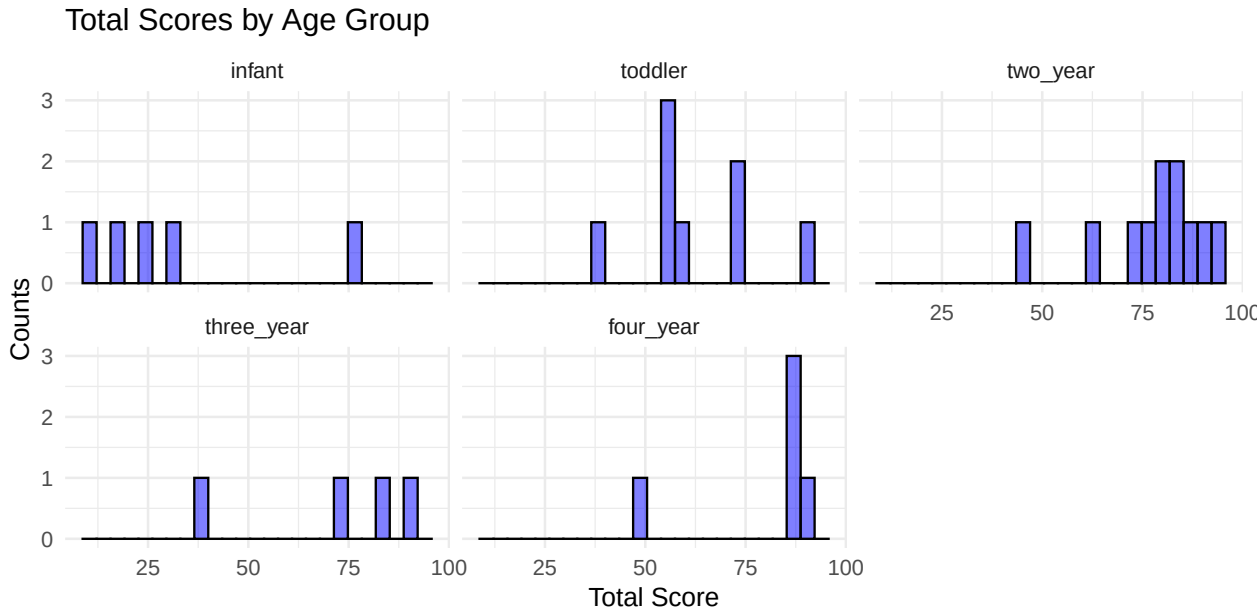


```
ggplot(UMCH, aes(x = AgeGroup,  y = LanguageDevelopment, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Language Development Score",
       title = "Number of Passes and Fails by Language Development Score")
```



- How does score vary by age group?

```
gf_histogram(~TotalScore, data=UMCH, fill="blue", color='black') %>%
  gf_labs(title="Total Scores by Age Group", x="Total Score", y="Counts") + facet_wrap(~AgeGroup)
```



Draft Models

Model Predicting PhysicalDevelopment

Response variable: PhysicalDevelopment
Predictor(s): LanguageDevelopment, TotalScore
Regression model: Linear regression

This model will show if there is a relationship between the Language Development and Physical Development of the child. We chose these two because all the age groups are graded on that category. In this first model, Physical development is the response variable. We also chose to add TotalScore to see if knowing how well the child did on the whole exam would help the prediction. This is a linear model because the response variable is continuous. We can use up to 2 predictors given the size of the data set (33/15). We will not add any interaction or random effect due to the size of our dataset. If it was larger, it would be interesting to explore the interaction between Language Development and Status. Perhaps whether or not a student passed affect the degree of which Language development can impact physical development. The same could be explored for the interaction between Language Development and Age group. But we may be able to get some insight into those with our prediction plots.

```
mod_phy <- lm(PhysicalDevelopment ~ TotalScore + LanguageDevelopment,
          data = UMCH)

summary(mod_phy)
```

```
##
## Call:
## lm(formula = PhysicalDevelopment ~ TotalScore + LanguageDevelopment,
##     data = UMCH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.1493  -2.2862   0.4821   4.6732   9.2016
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.2705     3.2460   0.083 0.934134
## TotalScore            0.6120     0.1181   5.181  1.4e-05 ***
## LanguageDevelopment  -0.7478     0.1964  -3.808 0.000645 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.793 on 30 degrees of freedom
## Multiple R-squared:  0.533,  Adjusted R-squared:  0.5019
## F-statistic: 17.12 on 2 and 30 DF,  p-value: 1.095e-05
```

Model Predicting LanguageDevelopment

Response variable: LanguageDevelopment
Predictor(s): PhysicalDevelopment, TotalScore
Regression model: Linear regression

This model will show if there is a relationship between the Language Development and Physical Development of the child. We chose these two because all the age groups are graded on that category. In this first model, Language development is the response variable. We also chose to add TotalScore to see if knowing how well the child did on the whole exam would help the prediction. This is a linear model because the response

variable is continuous. We can use up to 2 predictors given the size of the data set (33/15). We will not add any interaction or random effect due to the size of our dataset.

```
mod_lan <- lm(LanguageDevelopment ~ TotalScore + PhysicalDevelopment,
          data = UMCH)

summary(mod_lan)
```

```
##
## Call:
## lm(formula = LanguageDevelopment ~ TotalScore + PhysicalDevelopment,
##     data = UMCH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0883 -2.9101  0.9932  2.8127 10.3993
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -3.38171    2.40009  -1.409 0.169119
## TotalScore           0.64351    0.04005  16.069 2.77e-16 ***
## PhysicalDevelopment -0.43578    0.11443  -3.808 0.000645 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.423 on 30 degrees of freedom
## Multiple R-squared:  0.9079, Adjusted R-squared:  0.9018
## F-statistic: 147.9 on 2 and 30 DF,  p-value: 2.906e-16
```

Alternative Model

This is another alternative model we are thinking after we receive "Gender" information on the children next week.

Response variable: TotalScore Predictor(s): Age, Gender
Regression model: Linear regression

We chose TotalScore as our response variable and Age and Gender as our predictor variable because we want to know whether the Age and Gender have effect on the TotalScore. We chose linear regression model because it is used to show or predict the relationship between two variables where the response variable is continuous. We can use up to 2 predictors given the size of the data set (33/15). We will not add any interaction or random effect due to the size of our dataset.

— sample r code for alternative model —-
mod_alt <- lm(TotalScore ~ Age + Gender,
data = UMCH)
summary(mod_alt)