

UMCH

Cora Boyoung Jung, Enoch Mwesigwa, Jordan Severn

Import csv & Change data type

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

glimpse(UMCHdata)

## Rows: 36
## Columns: 22
## $ Filename                <chr> "failed_infant_1", "failed_infant_2...
## $ Birthdate              <date> 2020-05-26, 2020-06-03, 2020-03-05...
## $ AgeGroup               <chr> "infant", "infant", "infant", "infa...
## $ Age                    <int> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,...
## $ Gender                 <chr> "Male", "Female", "Female", "Female...
## $ PhysicalDevelopment    <dbl> 4.0, 4.0, 13.0, 23.0, 6.0, 24.0, 19...
## $ LanguageDevelopment   <dbl> 7.0, 5.0, 11.0, 32.0, 11.0, 36.0, 2...
## $ Adaptive_SelfHelp     <dbl> 3, 3, 4, 6, 3, 11, 8, 8, 8, 6, 7, 8...
## $ Adaptive_SocialEmotional <dbl> 3, 0, 5, 14, 5, 16, 12, 12, 12, 12,...
## $ AcademicAndCognitive  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ AcademicAndCognitive_Maths <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ AcademicAndCognitive_Literacy <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ PhysicalDevelopment_P <dbl> 14.29, 14.29, 46.43, 82.14, 21.43, ...
## $ LanguageDevelopment_P <dbl> 15.91, 11.36, 25.00, 72.73, 25.00, ...
## $ Adaptive_SelfHelp_P   <dbl> 25.00, 25.00, 33.33, 50.00, 25.00, ...
## $ Adaptive_SocialEmotional_P <dbl> 18.75, 0.00, 31.25, 87.50, 31.25, 1...
## $ AcademicAndCognitive_P <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ AcademicAndCognitive_Maths_P <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ AcademicAndCognitive_Literacy_P <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA,...
## $ TotalScore            <dbl> 17.0, 12.0, 33.0, 75.0, 25.0, 87.0,...
## $ Status                <chr> "failed", "failed", "failed", "pass...
## $ Examiner              <chr> "Sam McGowen", "Sam McGowen", "Sam ...
```

Tidying data

```
UMCH <- UMCHdata %>% mutate(Status = tolower(Status))
neworder <- c("infant", "toddler", "two_year", "three_year", "four_year")

UMCH <- arrange(transform(UMCH,
  AgeGroup=factor(AgeGroup, levels=neworder)), AgeGroup)
```

Statistics in Percentage

```
# group and get median of scores in each age group
UMCH_grouped <- UMCH %>%
  select(AgeGroup, PhysicalDevelopment, LanguageDevelopment, Adaptive_SelfHelp,
         Adaptive_SocialEmotional, AcademicAndCognitive, AcademicAndCognitive_Maths,
         AcademicAndCognitive_Literacy) %>%
  group_by(AgeGroup) %>%
  summarise_each(funs(round(mean(., na.rm = TRUE), digits=2)))

## Warning: `summarise_each()` is deprecated as of dplyr 0.7.0.
## Please use `across()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

## Warning: `funs()` is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

UMCH_grouped <- arrange(transform(UMCH_grouped,
                                AgeGroup=factor(AgeGroup,levels=neworder)),AgeGroup)

UMCH_grouped%>%
  knitr::kable()
```

AgeGroup	PhysicalDevelopment	LanguageDevelopment	Adaptive_SelfHelp	Adaptive_SocialEmotional	AcademicAndCognitive	AcademicAndCognitive_Maths	AcademicAndCognitive_Literacy
infant	12.33	17.00	5.00	7.17	NaN	NaN	NaN
toddler	15.89	25.94	7.67	12.00	NaN	NaN	NaN
two_year	24.09	38.05	NaN	NaN	15.64	NaN	NaN
three_year	21.75	35.25	NaN	NaN	15.00	NaN	NaN
four_year	7.17	42.83	NaN	NaN	6.67	13	6.83

Exploring data set

```
head(UMCH)

##      Filename  Birthdate AgeGroup Age Gender PhysicalDevelopment
## 1 failed_infant_1 2020-05-26  infant  0  Male                4
## 2 failed_infant_2 2020-06-03  infant  0 Female                4
## 3 failed_infant_3* 2020-03-05  infant  0 Female               13
## 4 passed_infant_1 2019-11-13  infant  0 Female               23
## 5 passed_infant_2 2020-07-01  infant  0  Male                 6
## 6 passed_infant_3 2019-11-30  infant  0  Male                24
```

```
## LanguageDevelopment Adaptive_SelfHelp Adaptive_SocialEmotional
## 1 7 3 3
## 2 5 3 0
## 3 11 4 5
## 4 32 6 14
## 5 11 3 5
## 6 36 11 16
## AcademicAndCognitive AcademicAndCognitive_Maths AcademicAndCognitive_Literacy
## 1 NA NA NA
## 2 NA NA NA
## 3 NA NA NA
## 4 NA NA NA
## 5 NA NA NA
## 6 NA NA NA
## PhysicalDevelopment_P LanguageDevelopment_P Adaptive_SelfHelp_P
## 1 14.29 15.91 25.00
## 2 14.29 11.36 25.00
## 3 46.43 25.00 33.33
## 4 82.14 72.73 50.00
## 5 21.43 25.00 25.00
## 6 85.71 81.82 91.67
## Adaptive_SocialEmotional_P AcademicAndCognitive_P
## 1 18.75 NA
## 2 0.00 NA
## 3 31.25 NA
## 4 87.50 NA
## 5 31.25 NA
## 6 100.00 NA
## AcademicAndCognitive_Maths_P AcademicAndCognitive_Literacy_P TotalScore
## 1 NA NA 17
## 2 NA NA 12
## 3 NA NA 33
## 4 NA NA 75
## 5 NA NA 25
## 6 NA NA 87
## Status Examiner
## 1 failed Sam McGowen
## 2 failed Sam McGowen
## 3 failed Sam McGowen
## 4 passed Melissa Swanson
## 5 passed Sam McGowen
## 6 passed Melissa Swanson
```

summary(UMCH)

```
## Filename Birthdate AgeGroup Age
## Length:36 Min. :2016-01-21 infant : 6 Min. :0.000
## Class :character 1st Qu.:2017-06-26 toddler : 9 1st Qu.:1.000
## Mode :character Median :2018-04-10 two_year :11 Median :2.000
## Mean :2018-04-29 three_year: 4 Mean :1.861
## 3rd Qu.:2019-03-31 four_year : 6 3rd Qu.:3.000
## Max. :2020-07-01 Max. :4.000
##
## Gender PhysicalDevelopment LanguageDevelopment Adaptive_SelfHelp
## Length:36 Min. : 2.00 Min. : 5.00 Min. : 3.0
```

```

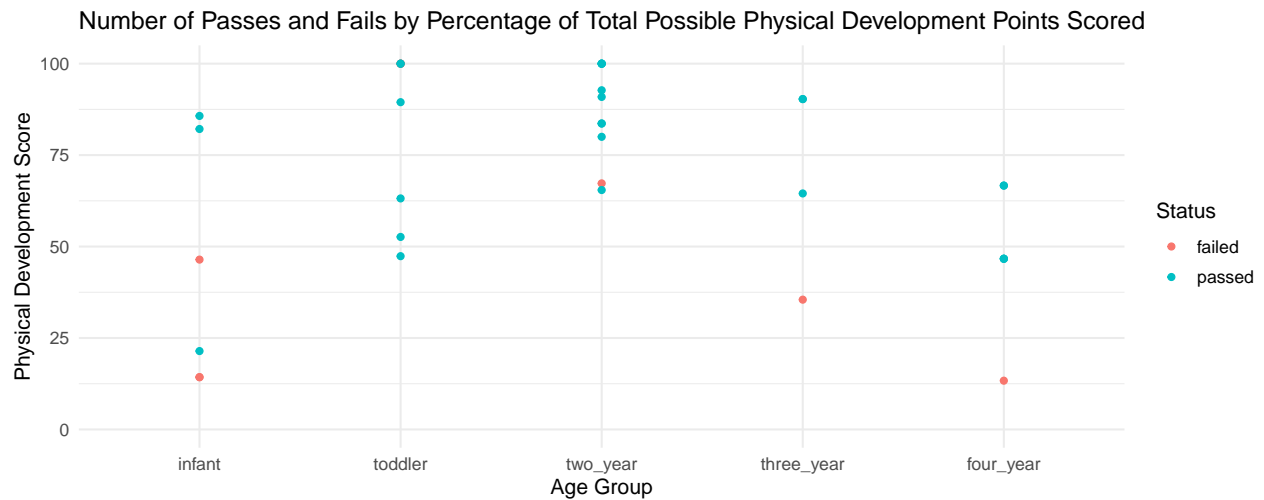
## Class :character      1st Qu.:10.00      1st Qu.:20.75      1st Qu.: 5.0
## Mode :character      Median :19.00      Median :35.75      Median : 8.0
##                               Mean :17.00      Mean :32.00      Mean : 6.6
##                               3rd Qu.:23.25      3rd Qu.:43.25      3rd Qu.: 8.0
##                               Max. :28.00      Max. :51.00      Max. :11.0
##                               NA's :21
## Adaptive_SocialEmotional AcademicAndCognitive AcademicAndCognitive_Maths
## Min. : 0.00      Min. : 2.50      Min. : 5.000
## 1st Qu.: 8.00      1st Qu.: 7.50      1st Qu.: 9.375
## Median :12.00      Median :12.00      Median :15.250
## Mean :10.07      Mean :12.95      Mean :13.000
## 3rd Qu.:12.00      3rd Qu.:18.00      3rd Qu.:16.625
## Max. :16.00      Max. :21.50      Max. :18.000
## NA's :21      NA's :15      NA's :30
## AcademicAndCognitive_Literacy PhysicalDevelopment_P LanguageDevelopment_P
## Min. : 0.000      Min. : 13.33      Min. : 11.36
## 1st Qu.: 3.750      1st Qu.: 47.20      1st Qu.: 36.84
## Median : 9.500      Median : 81.07      Median : 71.44
## Mean : 6.833      Mean : 70.77      Mean : 63.26
## 3rd Qu.:10.000      3rd Qu.: 94.55      3rd Qu.: 84.84
## Max. :10.000      Max. :100.00      Max. :100.00
## NA's :30
## Adaptive_SelfHelp_P Adaptive_SocialEmotional_P AcademicAndCognitive_P
## Min. :25.00      Min. : 0.00      Min. : 19.05
## 1st Qu.:41.66      1st Qu.: 57.94      1st Qu.: 55.81
## Median :88.89      Median : 92.31      Median : 75.00
## Mean :67.78      Mean : 73.30      Mean : 70.75
## 3rd Qu.:88.89      3rd Qu.: 92.31      3rd Qu.: 85.71
## Max. :91.67      Max. :100.00      Max. :100.00
## NA's :21      NA's :21      NA's :15
## AcademicAndCognitive_Maths_P AcademicAndCognitive_Literacy_P TotalScore
## Min. : 29.41      Min. : 0.00      Min. :12.00
## 1st Qu.: 55.15      1st Qu.: 37.50      1st Qu.:54.00
## Median : 89.71      Median : 95.00      Median :73.75
## Mean : 76.47      Mean : 68.33      Mean :66.81
## 3rd Qu.: 97.80      3rd Qu.:100.00      3rd Qu.:86.00
## Max. :105.88      Max. :100.00      Max. :95.50
## NA's :30      NA's :30
## Status      Examiner
## Length:36      Length:36
## Class :character      Class :character
## Mode :character      Mode :character
##
##
##
##

```

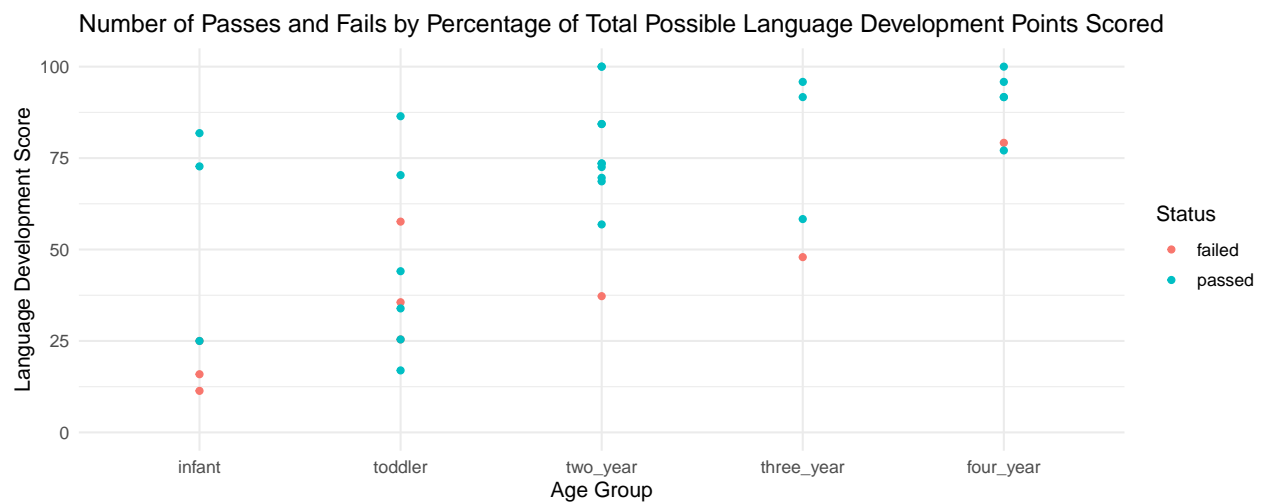
Graphics

- 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_P, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Physical Development Score",
       title = "Number of Passes and Fails by Percentage of Total Possible Physical Development Points Scored",
       expand_limits(y = 0))
```

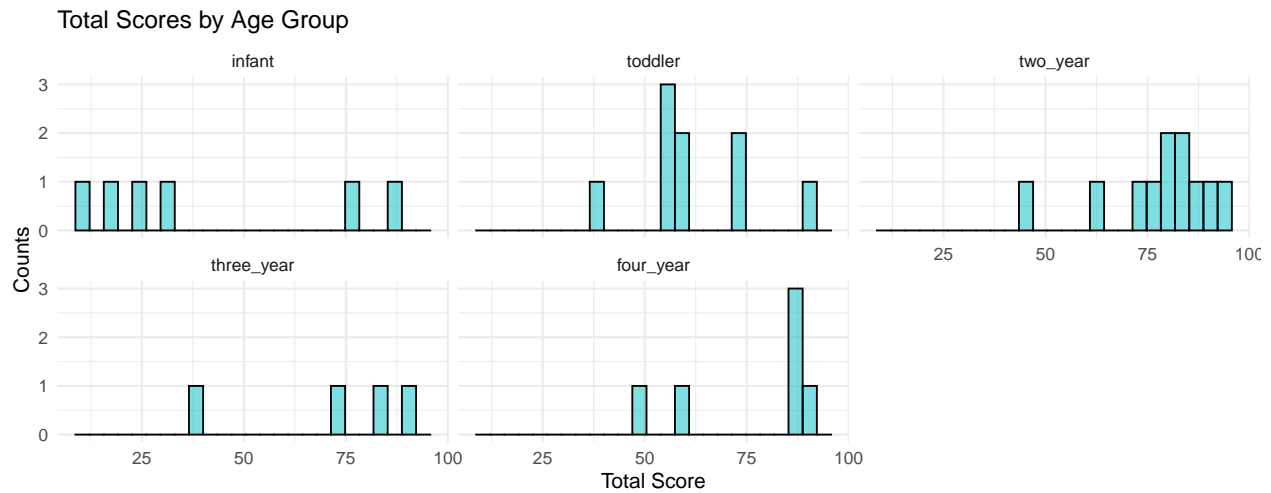


```
ggplot(UMCH, aes(x = AgeGroup, y = LanguageDevelopment_P, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Language Development Score",
       title = "Number of Passes and Fails by Percentage of Total Possible Language Development Points Scored",
       expand_limits(y = 0))
```



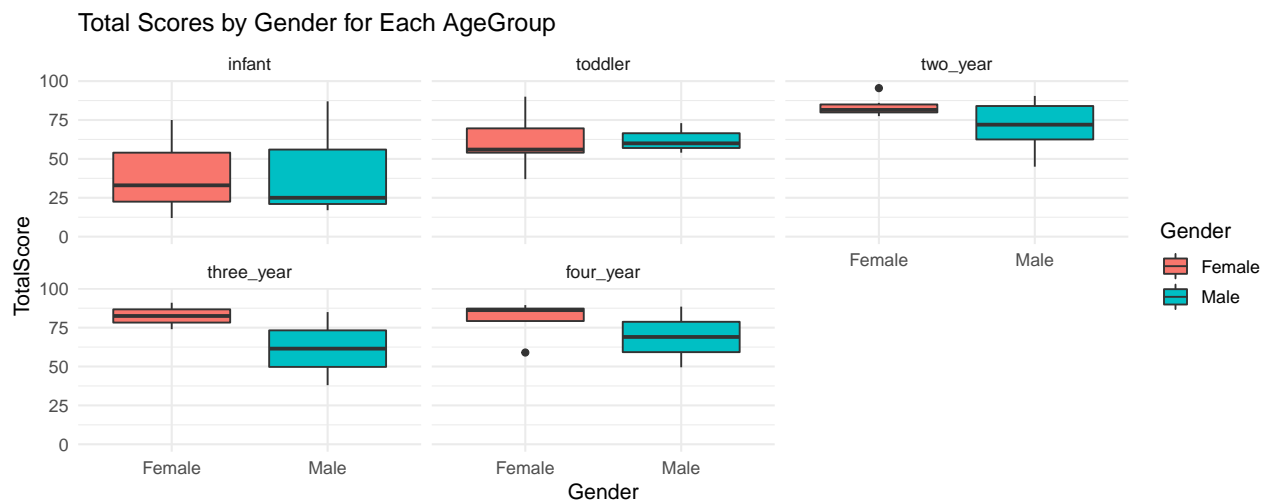
- How does score vary by age group?

```
gf_histogram(~TotalScore, data=UMCH, fill='#00BFC4', color='black') %>%
  gf_labs(title="Total Scores by Age Group", x="Total Score", y="Counts") + facet_wrap(~AgeGroup)
```



- How does score vary by gender?

```
ggplot(UMCH, aes(x = Gender, y = TotalScore, fill=Gender)) +
  geom_boxplot() +
  facet_wrap(~AgeGroup) +
  labs(title="Total Scores by Gender for Each AgeGroup", x="Gender") +
  expand_limits(y = 0)
```



Statistics in Raw Score

```
UMCH_long <- UMCH %>%
  subset(select = -c(Filename, Birthdate, Age, Gender, TotalScore, Status, Examiner, Adaptive_SelfHelp, A
  pivot_longer(!AgeGroup, names_to = "Domain", values_to = "Score") %>%
  na.omit() %>%
  mutate(AgeGroup = as.factor(AgeGroup),
         Domain = as.factor(Domain),
         Score = as.numeric(Score))

UMCH_stat <- UMCH_long %>%
  dplyr::group_by(AgeGroup, Domain) %>%
  dplyr::summarise(Min=min(Score),
```

```

Max=max(Score),
Mean=mean(Score),
Median=median(Score))

```

```

UMCH_stat %>%
  knitr::kable()

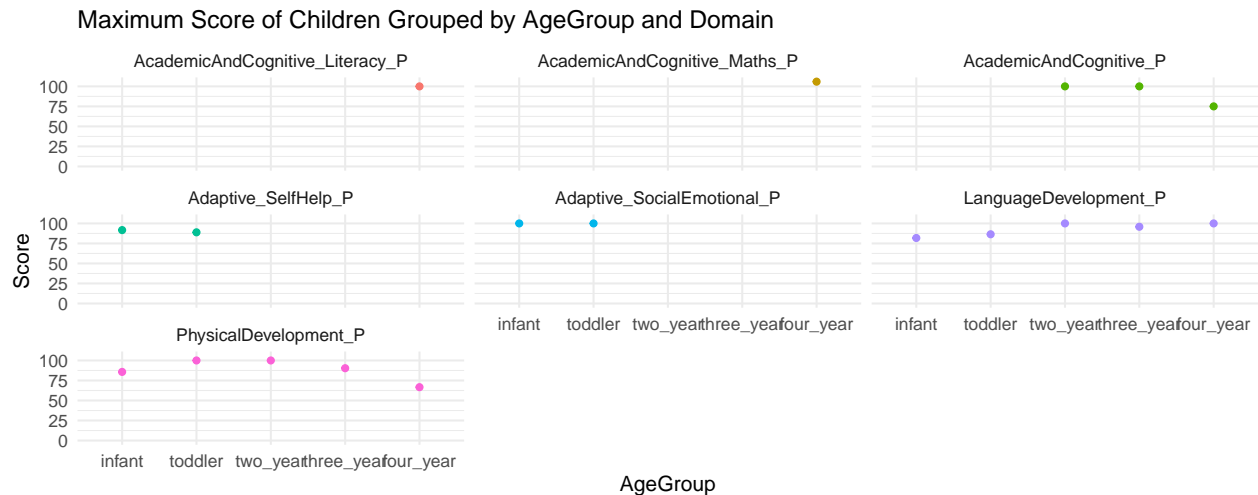
```

AgeGroup	Domain	Min	Max	Mean	Median
infant	Adaptive_SelfHelp_P	25.00	91.67	41.66667	29.165
infant	Adaptive_SocialEmotional_P	0.00	100.00	44.79167	31.250
infant	LanguageDevelopment_P	11.36	81.82	38.63667	25.000
infant	PhysicalDevelopment_P	14.29	85.71	44.04833	33.930
toddler	Adaptive_SelfHelp_P	66.67	88.89	85.18667	88.890
toddler	Adaptive_SocialEmotional_P	84.62	100.00	92.31000	92.310
toddler	LanguageDevelopment_P	16.95	86.44	43.97333	35.590
toddler	PhysicalDevelopment_P	47.37	100.00	83.62556	100.000
two_year	AcademicAndCognitive_P	34.88	100.00	72.72727	67.440
two_year	LanguageDevelopment_P	37.25	100.00	74.59818	73.530
two_year	PhysicalDevelopment_P	65.45	100.00	87.60364	90.910
three_year	AcademicAndCognitive_P	19.05	100.00	71.42750	83.330
three_year	LanguageDevelopment_P	47.92	95.83	73.43750	75.000
three_year	PhysicalDevelopment_P	35.48	90.32	70.16000	77.420
four_year	AcademicAndCognitive_Literacy_P	0.00	100.00	68.33333	95.000
four_year	AcademicAndCognitive_Maths_P	29.41	105.88	76.47167	89.710
four_year	AcademicAndCognitive_P	25.00	75.00	66.66667	75.000
four_year	LanguageDevelopment_P	77.08	100.00	89.23667	91.670
four_year	PhysicalDevelopment_P	13.33	66.67	47.78000	46.670

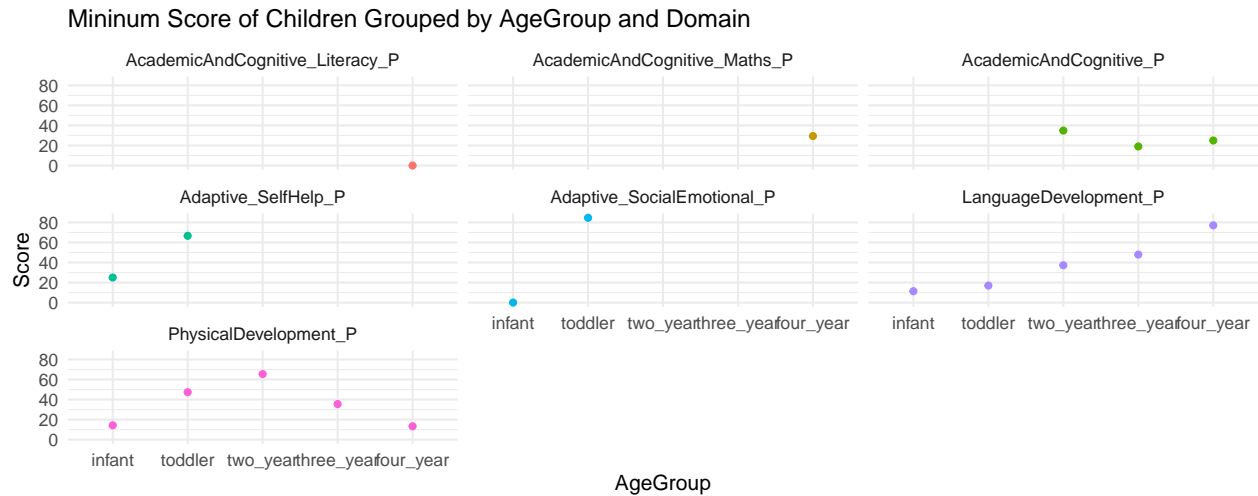
```

ggplot(UMCH_stat, aes(x = AgeGroup, y = Max, color=Domain)) +
  geom_point() +
  facet_wrap(~Domain) +
  labs(title="Maximum Score of Children Grouped by AgeGroup and Domain", y="Score") +
  theme(legend.position = "none") +
  expand_limits(y = 0)

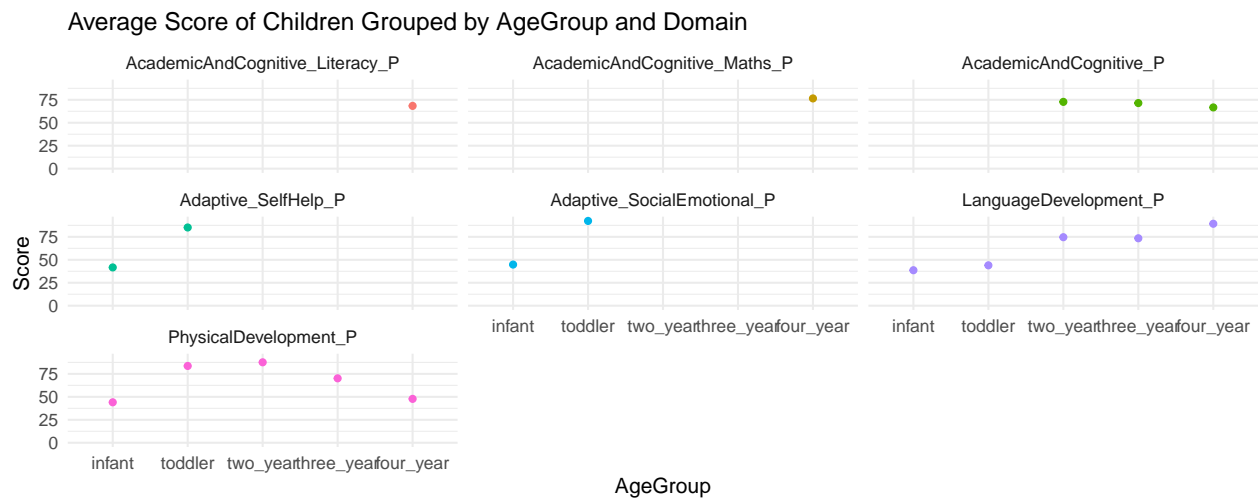
```



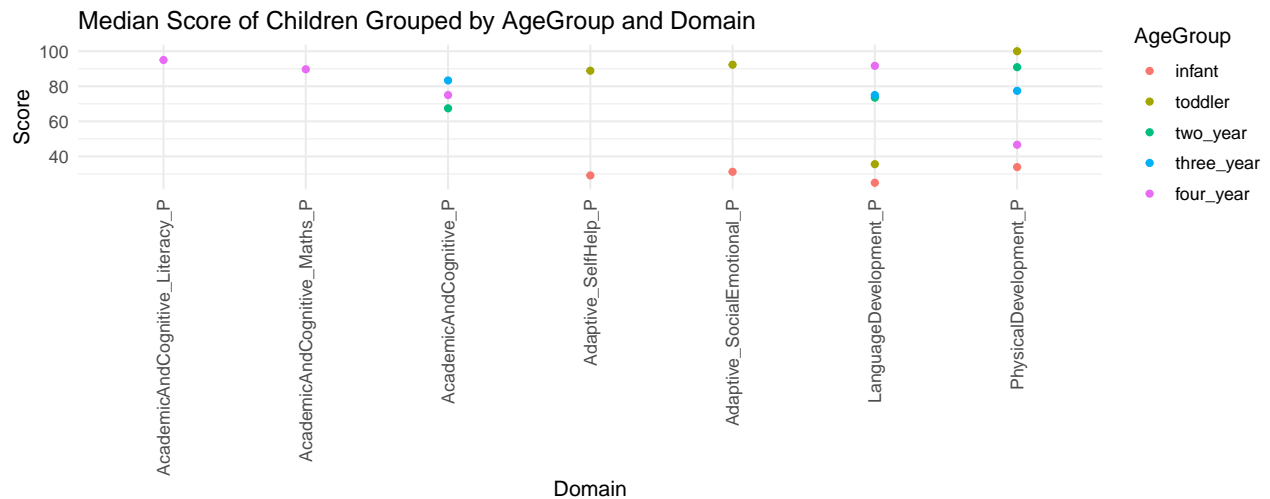
```
ggplot(UMCH_stat, aes(x = AgeGroup, y = Min, color=Domain)) +
  geom_point() +
  facet_wrap(~Domain) +
  labs(title="Minimum Score of Children Grouped by AgeGroup and Domain", y="Score") +
  theme(legend.position = "none") +
  expand_limits(y = 0)
```



```
ggplot(UMCH_stat, aes(x = AgeGroup, y = Mean, color=Domain)) +
  geom_point() +
  facet_wrap(~Domain) +
  labs(title="Average Score of Children Grouped by AgeGroup and Domain", y="Score") +
  theme(legend.position = "none") +
  expand_limits(y = 0)
```



```
ggplot(UMCH_stat, aes(x = Domain, y = Median, color=AgeGroup)) +
  geom_point() +
  labs(title="Median Score of Children Grouped by AgeGroup and Domain", y="Score") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

```
expand_limits(y = 0)
```

```
## mapping: y = ~y
## geom_blank: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

TotalScore Model

We used linear regression model because it is used to show/predict the relationship between variables where the response variable is continuous. We can use up to 2 predictors given the size of the data (36/15). We will not add any interaction or random effect due to the size of the dataset. Our predictors will be Age and Gender because all other variables contributes to the Total Score which does not make logical sense to include those.

Response variable: TotalScore Predictor(s): Age, Gender

Regression model: Linear regression

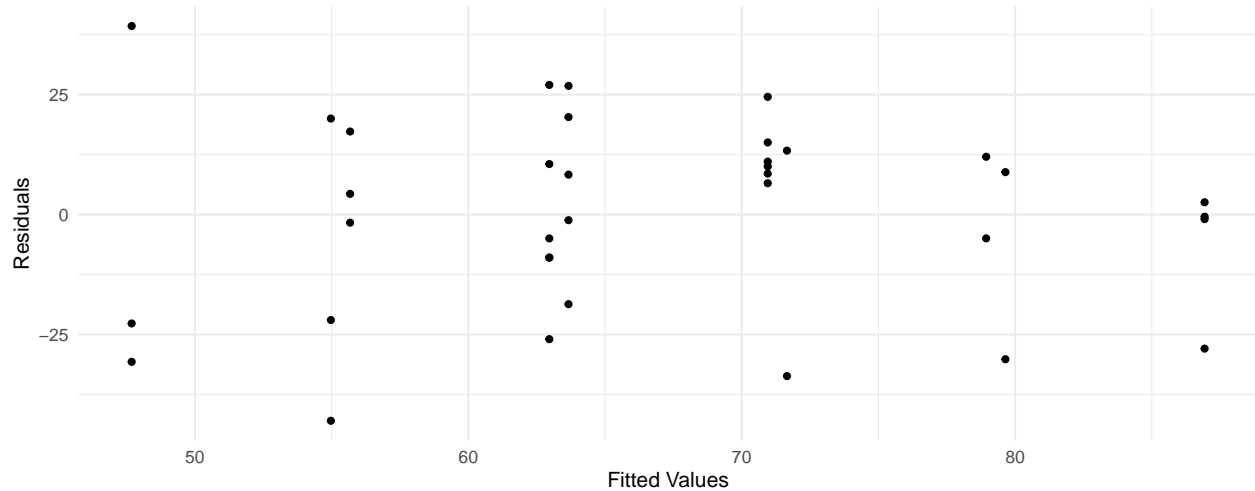
```
mod_total <- lm(TotalScore ~ Age + Gender,
  data = UMCH)
summary(mod_total)
```

```
##
## Call:
## lm(formula = TotalScore ~ Age + Gender, data = UMCH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42.971 -11.387   3.447  12.380  39.316
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   54.971      6.700   8.205 1.79e-09 ***
## Age           7.990      2.629   3.039 0.00461 **
## GenderMale    -7.287      6.902  -1.056 0.29872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.4 on 33 degrees of freedom
```

```
## Multiple R-squared:  0.2436, Adjusted R-squared:  0.1977
## F-statistic: 5.313 on 2 and 33 DF,  p-value: 0.009995
```

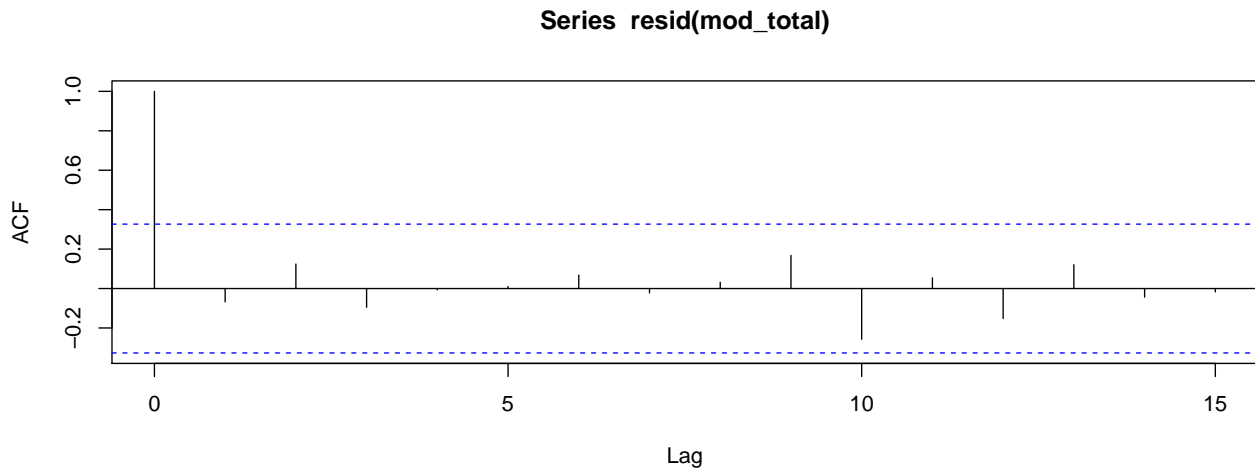
Model Assessment

```
gf_point(resid(mod_total) ~ fitted(mod_total)) %>%
  gf_labs(x = 'Fitted Values', y = 'Residuals')
```



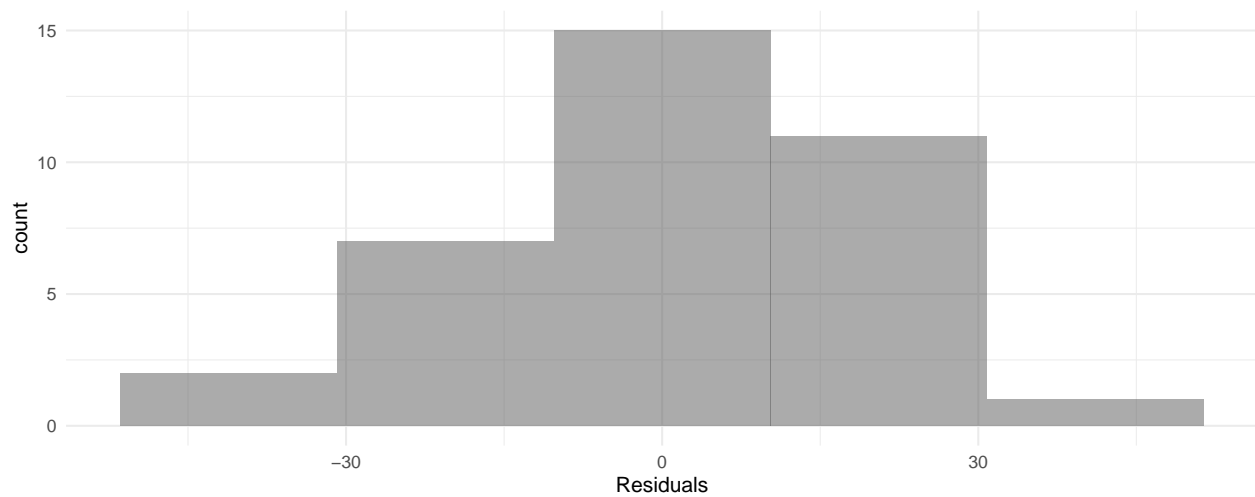
The fitted values vs. the residual plot is used for linearity and error variance. There is no trumpeting present so it passes the error variance condition. However, there is some clustering present and the data points make vertical lines so linearity is a little more iffy on whether or not it passes.

```
acf(resid(mod_total))
```



Looking at the acf plot, the plot passes the independence condition. All the lines are within the dotted lines and the lag at 0 is normal.

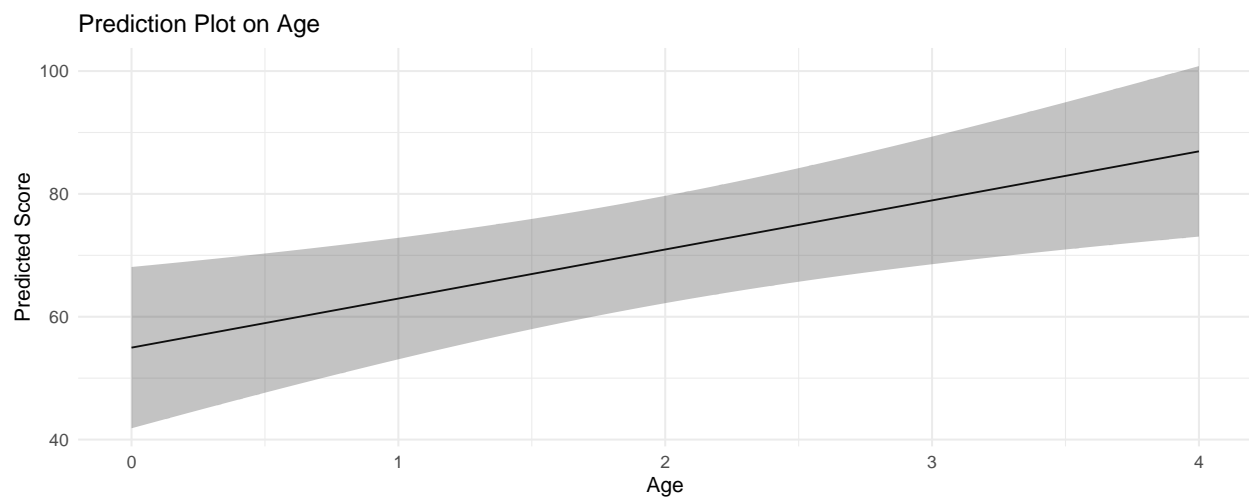
```
gf_histogram(~resid(mod_total), bins = 5) %>%
  gf_labs(x = "Residuals")
```



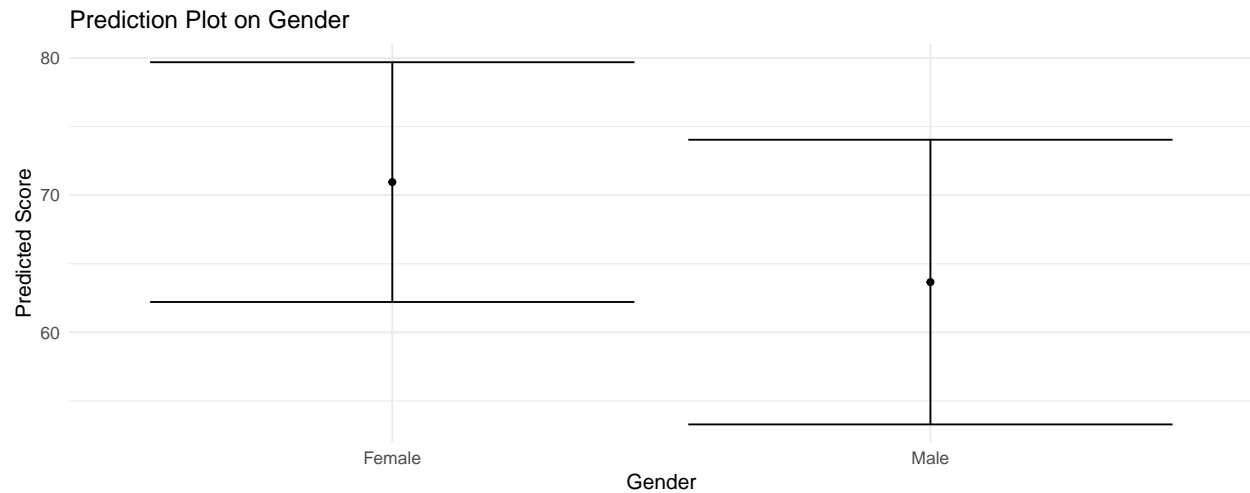
Since the data is limited, the bins need to be limited. The histogram passes the normality condition.

Prediction Plot

```
pred_plot(mod_total, 'Age') %>%
  gf_labs(y = 'Predicted Score',
          title = 'Prediction Plot on Age')
```



```
pred_plot(mod_total, 'Gender') %>%
  gf_labs(y = 'Predicted Score',
          title = 'Prediction Plot on Gender')
```



According to the prediction plot for Age, there seems to be a positive trend between the score and age where as age increases, the score also increases. This makes sense because the “pass” score for each age group differs according to the age with lower age having the lowest pass score.

According to the prediction plot for Gender, it looks like the total score is a little higher for female compared to male. However, the overlapping range is high and the number of children is not the same (21 females and 15 males), so it brings down the reliability score of this data but it is worth noting.