

Sarah & Joe Present: Does time spent doing homework in adolescence increasing reading and math achievement?

Packages

```
library(here)
library(purrr)
library(dplyr)
library(tidyr)
library(haven)
library(forcats)
library(ggplot2)
library(stringr)
library(sjlabelled)
library(ggcorrplot)
```

Read in data and filter out all those who were not in the wave

```
data <- read_dta(here("class data.dta"))

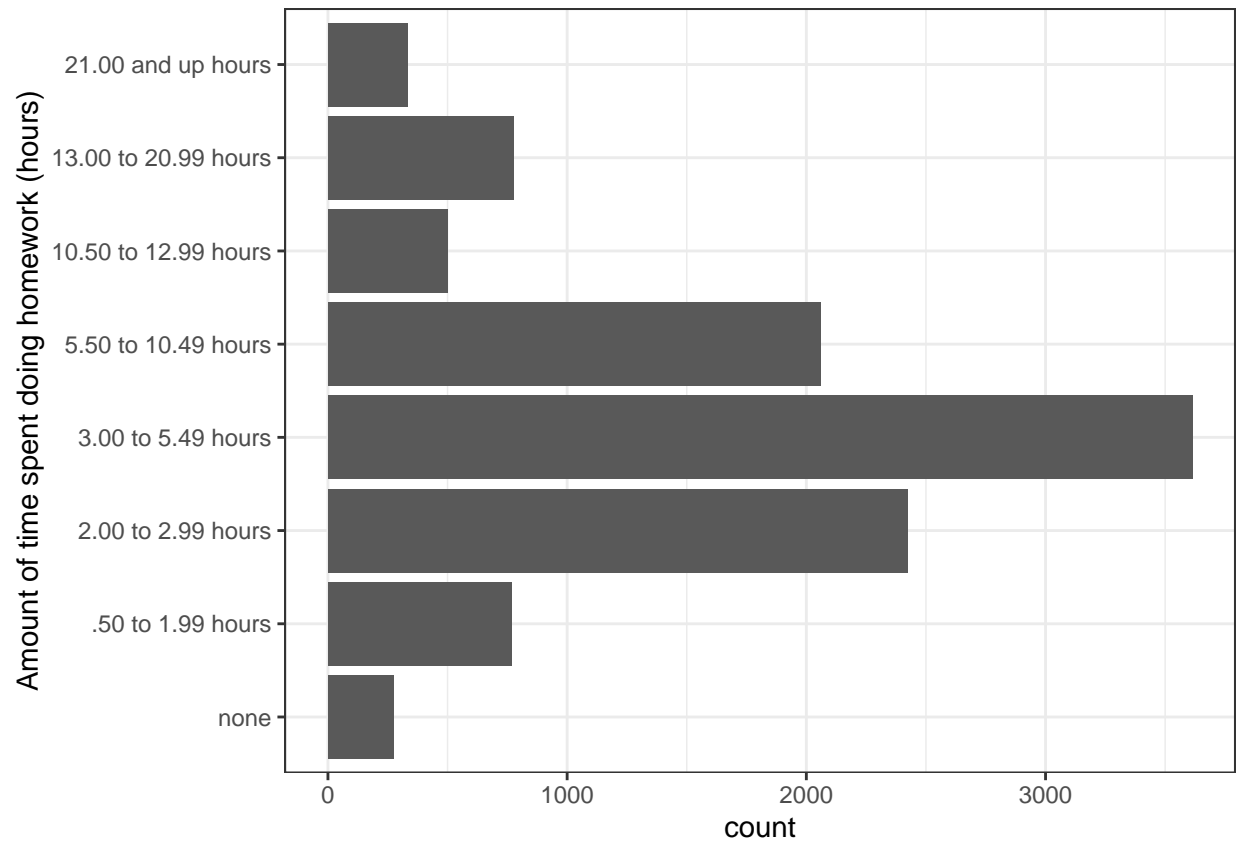
baseYearCohort <-
  data %>%
  select(matches("by|sex|race")) %>%
  filter(byhomewk != 99)
```

Number of students removed due to not being in the wave: 1397

Fraction of total students removed: 0.1150362

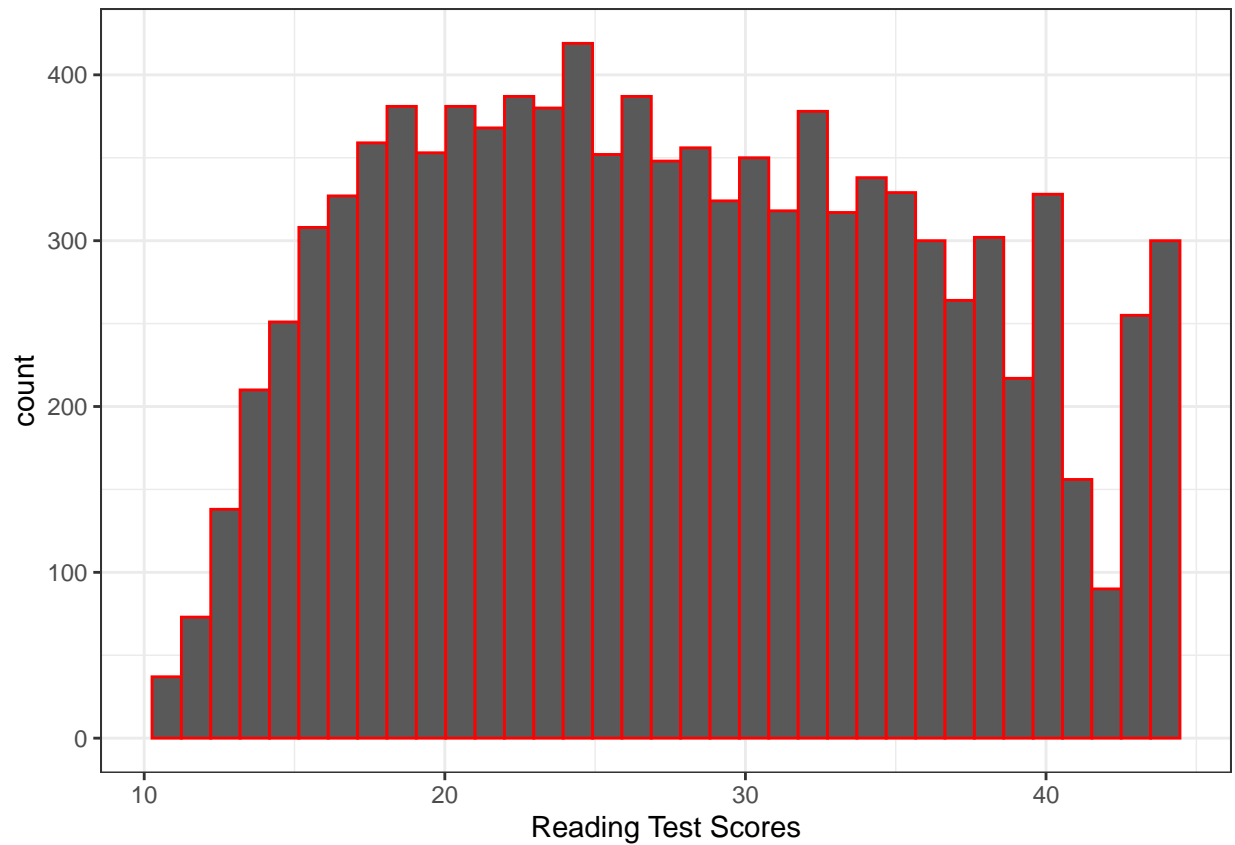
Get a sense for the data

```
ggplot(baseYearCohort, aes(x = as_label(byhomewk))) +
  geom_bar() +
  theme_bw() +
  coord_flip() +
  labs(x = "Amount of time spent doing homework (hours)")
```



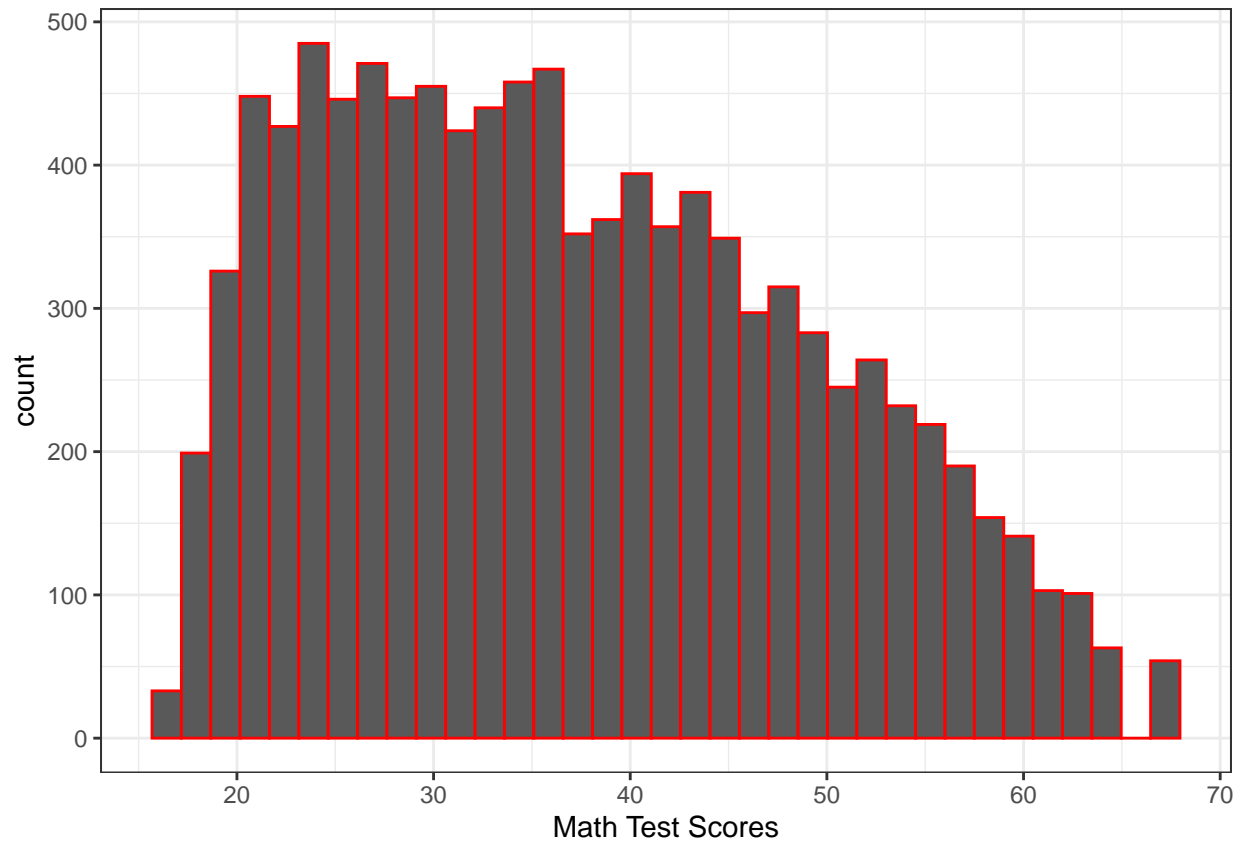
```
ggplot(baseYearCohort, aes(x = as.numeric(by2xrirr))) +  
  geom_histogram(bins = 35, color = "red") +  
  labs(x = "Reading Test Scores") +  
  theme_bw()
```

```
## Warning: Removed 366 rows containing non-finite values (stat_bin).
```



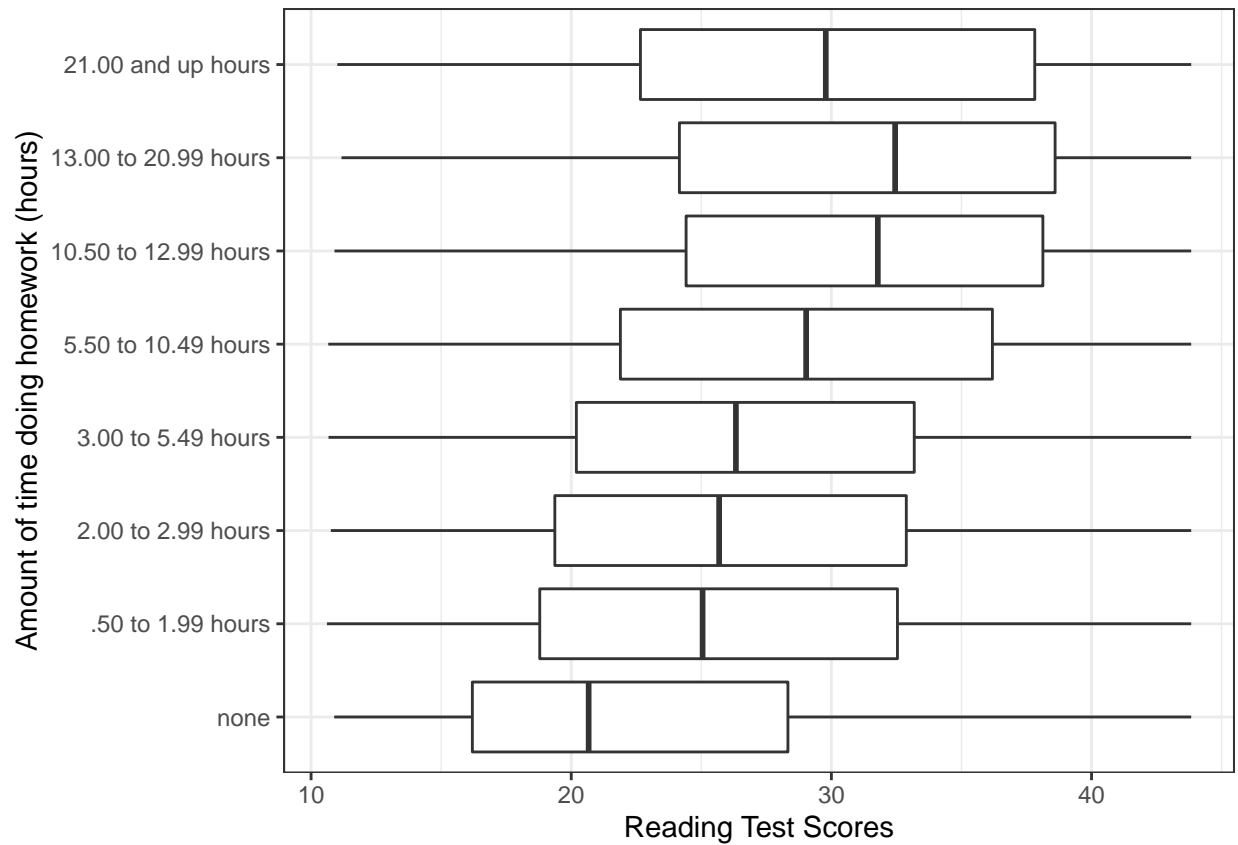
```
ggplot(baseYearCohort, aes(x = as.numeric(by2xmirr))) +  
  geom_histogram(bins = 35, color = "red") +  
  labs(x = "Math Test Scores") +  
  theme_bw()
```

```
## Warning: Removed 365 rows containing non-finite values (stat_bin).
```



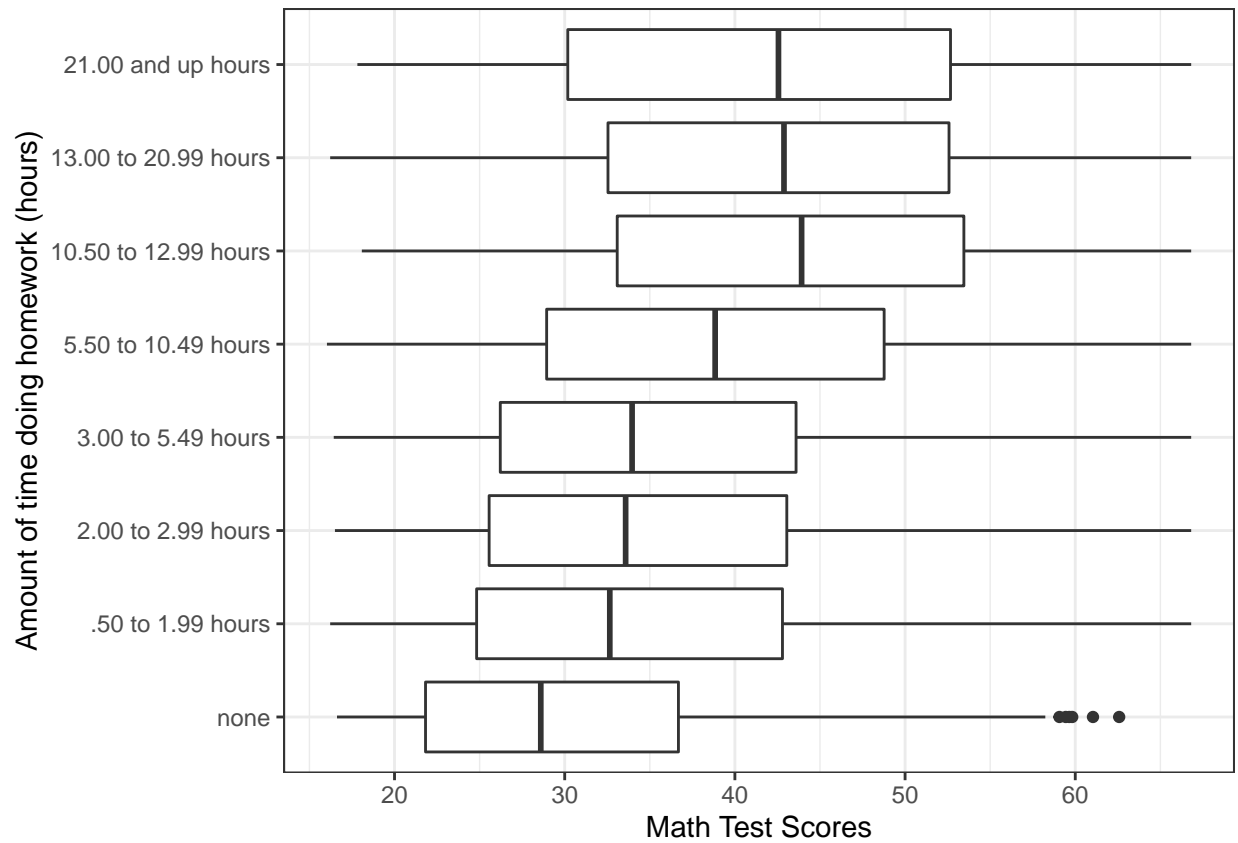
```
ggplot(baseYearCohort, aes(x = as_label(byhomewk), y = as.numeric(by2xrirr))) +
  geom_boxplot() +
  theme_bw() +
  coord_flip() +
  labs(x = "Amount of time doing homework (hours)", y = "Reading Test Scores")
```

```
## Warning: Removed 366 rows containing non-finite values (stat_boxplot).
```



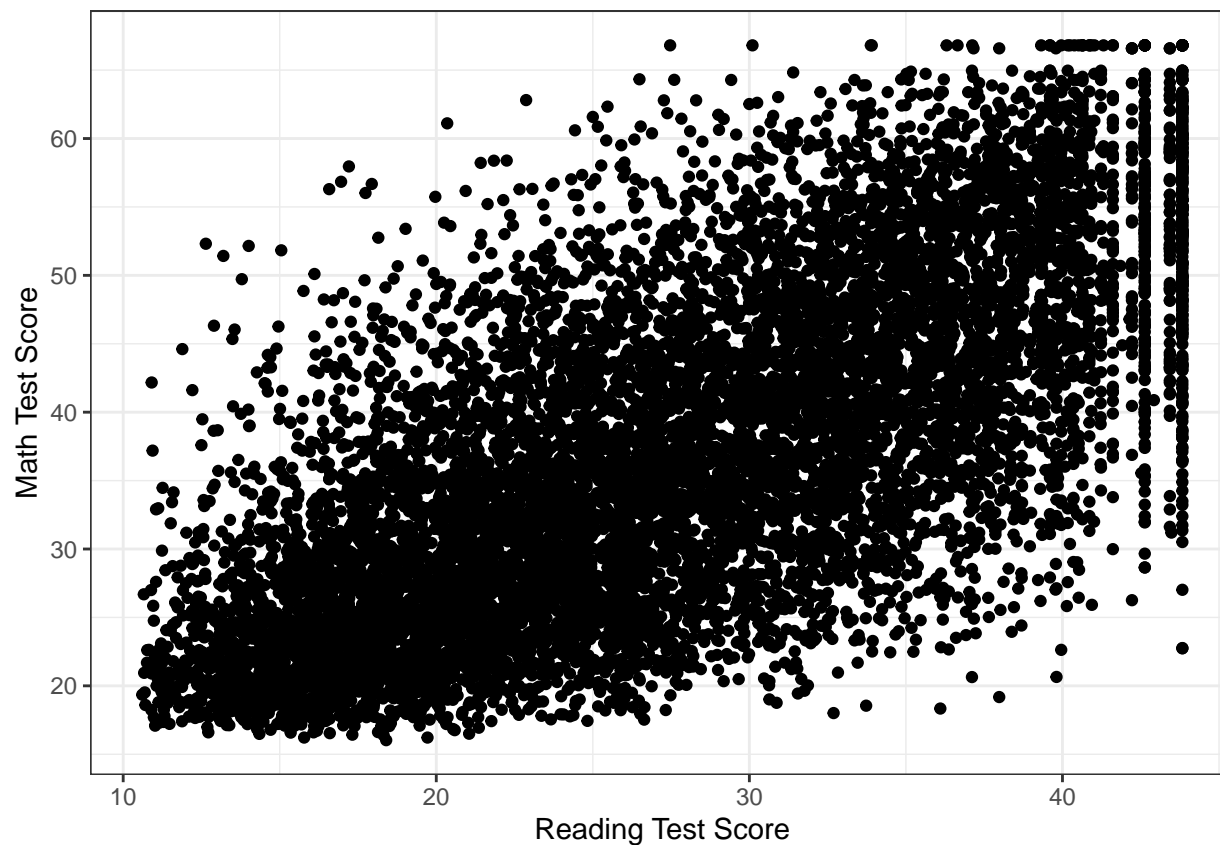
```
ggplot(baseYearCohort, aes(x = as_label(byhomewk), y = as.numeric(by2xmirr))) +
  geom_boxplot() +
  theme_bw() +
  coord_flip() +
  labs(x = "Amount of time doing homework (hours)", y = "Math Test Scores")
```

Warning: Removed 365 rows containing non-finite values (stat_boxplot).



```
ggplot(baseYearCohort, aes(x = as.numeric(by2xrirr), y = as.numeric(by2xmirr))) +
  labs(x = "Reading Test Score", y = "Math Test Score") +
  geom_point() +
  theme_bw()
```

Warning: Removed 380 rows containing missing values (geom_point).



```
cor(baseYearCohort$by2xrirr, baseYearCohort$by2xmirr, use = "complete.obs")
```

```
## [1] 0.7032982
```

Let's run some basic regressions

```
baseYearCohortWide <-
  baseYearCohort %>%
  mutate(byhomewk_label = as_label(byhomewk),
         byhomewk = as.numeric(as.logical(as.numeric(byhomewk)))) %>%
  pivot_wider(names_from = byhomewk_label,
             values_from = byhomewk,
             values_fill = F)

readingTestRegression <-
  lm(by2xrirr ~ `.50 to 1.99 hours` + `.2.00 to 2.99 hours` +
      `.3.00 to 5.49 hours` + `.5.50 to 10.49 hours` +
      `.10.50 to 12.99 hours` + `.13.00 to 20.99 hours` +
      `.21.00 and up hours`,
     data = baseYearCohortWide)
summary(readingTestRegression)
```

```
##
```

```
## Call:
```

```
## lm(formula = by2xrirr ~ `.50 to 1.99 hours` + `.2.00 to 2.99 hours` +
```

```
##      `.3.00 to 5.49 hours` + `.5.50 to 10.49 hours` + `.10.50 to 12.99 hours` +
```

```
##      `13.00 to 20.99 hours` + `21.00 and up hours`, data = baseYearCohortWide)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -20.0096  -6.8776  -0.4125   6.7024  20.9657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      22.8643     0.5196  44.001 < 2e-16 ***
## `.50 to 1.99 hours`    3.0414     0.6050   5.027 5.07e-07 ***
## `.2.00 to 2.99 hours`    3.5682     0.5482   6.508 7.95e-11 ***
## `.3.00 to 5.49 hours`    4.0433     0.5388   7.504 6.70e-14 ***
## `.5.50 to 10.49 hours`    6.0781     0.5534  10.984 < 2e-16 ***
## `.10.50 to 12.99 hours`    8.0453     0.6474  12.428 < 2e-16 ***
## `.13.00 to 20.99 hours`    8.2398     0.6058  13.602 < 2e-16 ***
## `.21.00 and up hours`    6.8078     0.7077   9.620 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.459 on 10373 degrees of freedom
## (366 observations deleted due to missingness)
## Multiple R-squared:  0.04154,    Adjusted R-squared:  0.0409
## F-statistic: 64.23 on 7 and 10373 DF,  p-value: < 2.2e-16
mathTestRegression <-
  lm(by2xmirr ~ `.50 to 1.99 hours` + `.2.00 to 2.99 hours` +
      `.3.00 to 5.49 hours` + `.5.50 to 10.49 hours` +
      `.10.50 to 12.99 hours` + `.13.00 to 20.99 hours` +
      `.21.00 and up hours`,
      data = baseYearCohortWide)
summary(mathTestRegression)

##
## Call:
## lm(formula = by2xmirr ~ `.50 to 1.99 hours` + `.2.00 to 2.99 hours` +
##      `.3.00 to 5.49 hours` + `.5.50 to 10.49 hours` + `.10.50 to 12.99 hours` +
##      `.13.00 to 20.99 hours` + `.21.00 and up hours`, data = baseYearCohortWide)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -26.349  -9.662  -1.080   8.684  32.325
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      30.8433     0.7170  43.019 < 2e-16 ***
## `.50 to 1.99 hours`    3.6421     0.8355   4.359 1.32e-05 ***
## `.2.00 to 2.99 hours`    4.1789     0.7566   5.523 3.41e-08 ***
## `.3.00 to 5.49 hours`    4.7171     0.7435   6.344 2.33e-10 ***
## `.5.50 to 10.49 hours`    8.3756     0.7636  10.968 < 2e-16 ***
## `.10.50 to 12.99 hours`   12.0024     0.8942  13.423 < 2e-16 ***
## `.13.00 to 20.99 hours`   11.7257     0.8366  14.016 < 2e-16 ***
## `.21.00 and up hours`   11.0460     0.9766  11.311 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```



```
## Residual standard error: 11.69 on 10374 degrees of freedom
## (365 observations deleted due to missingness)
## Multiple R-squared:  0.05817,    Adjusted R-squared:  0.05753
## F-statistic: 91.52 on 7 and 10374 DF,  p-value: < 2.2e-16
```