Final Project

Theresa Szczepanski

12/12/2022

Table of contents

```
library(tidyverse)
library(ggplot2)
library(lubridate)
library(readxl)
library(hrbrthemes)
library(viridis)
library(ggpubr)
library(purrr)
library(plotly)
knitr::opts_chunk$set(echo = TRUE, warning=FALSE, message=FALSE)
```

Introduction

Massachusetts Comprehensive Assessment System (MCAS) tests were introduced as part of the Massachusetts Education Reform Act in 1993 with the goal of providing all students with the skills and knowledge to thrive in a "complex and changing society" (Papay et. al, 2020 pp, 1). The MCAS tests are a significant tool for educational equity. Scores on the Grade 10 Math MCAS test "predict longer-term educational attainments and labor market success, above and beyond typical markers of student advantage. For example, among demographically similar students who attended the same high school and have the same level of ultimate educational attainment, those with higher MCAS mathematics scores go on to have much higher average earnings than those with lower scores." (Papay et. al, 2020 pp 7-10)

In this report, I will analyze the Spring 2022 MCAS Results for students completing the High School Introductory Physics MCAS at Rising Tide Charter Public School.

The `MCAS_2022` data frame contains performance results from 495 students from Rising Tide on the Spring 2022 Massachusetts Comprehensive Assessment System (MCAS) tests.

For each student, there are values reported for 256 different variables which consist of information from four broad categories

- *Demographic characteristics* of the students themselves (e.g., race, gender, date of birth, town, grade level, years in school, years in Massachusetts, and low income, title1, IEP, 504, and EL status).
- *Key assessment features* including subject, test format, and accommodations provided
- *Performance metrics*: This includes a student's score on individual item strands, e.g.,`sitem1-sitem42`.

See the `MCAS_2022` data frame summary and **codebook** in the **appendix** for further details.

The second data set, `SG9_Item`, is 42×9 and consists of 9 variables with information pertaining to the 42 questions on the 2022 HS Introductory Physics Item Report. The variables can be broken down into 2 categories:

Details about the content of a given test item:

This includes the content `Reporting Category` (MF (motion and forces) WA (waves), and EN (energy), the `Standard` from the 2016 STE Massachusetts Curriculum Framework, the `Item Description` providing the details of what specifically was asked of students, and the points available for a given question, `item Possible Points`.

Summary Performance Metrics:

- For each item, the state reports the percentage of points earned by students at Rising Tide, `RT Percent Points`, the percentage of available points earned by students in the state, `State Percent Points`, and the difference between the percentage of points earned by Rising Tide students and the percentage of points earned by students in the state, `RT-State Diff`.
- Lastly, `SG9_CU306Dis` and `SG9_CU306NonDis` are 3×5 dataframes consisting of summary performance data by `Reporting Category` for students with disabilities and without disabilities; most importantly including `RT Percent Points` and `State Percent Points`.

When considering our student performance data, we hope to address the following broad questions:

- What adjustments (if any) should be made at the Tier 1 level, i.e., curricular adjustments for all students?
- What would be the most beneficial areas of focus for a targeted intervention course for students struggling to meet or exceed performance expectations?

- Are there notable differences in student performance for students with and without disabilities?

Function Library

To read in, tidy, and join our data frames for each content area we will use the following functions. This still needs to be made more adaptable to include ELA.

```
#Item analysis Read in Function

#subject must be: "math", "ela", or "science"
read_item<-function(sheet_name, subject, grade){
  subject_item<-case_when(
    subject == "science"~"sitem",
    subject == "math"~"mitem",
    subject == "ela"~"eitem"
  )
  if(subject == "science"){
  read_excel("_data/2022MCASDepartmentalAnalysis.xlsx", sheet = sheet_name,
             skip = 1, col_names= c(subject_item, "Type", "Reporting Category", "Standard"
                  select(!contains("delete"))%>%
                  filter(!str_detect(sitem,"Legend|legend"))%>%
    mutate(sitem= as.character(sitem))%>%
    separate(c(1), c("sitem", "delete"))%>%
  select(!contains("delete"))%>%
   mutate(sitem =
            str_c(subject_item, sitem))
   }
  else if(subject == "math" && grade < 10){
     read_excel("_data/2022MCASDepartmentalAnalysis.xlsx", sheet = sheet_name,
              skip = 1, col_names= c(subject_item, "Type", "Reporting Category", "Standard
                   select(!contains("delete"))%>%
                   filter(!str_detect(mitem,"Legend|legend"))%>%
     mutate(mitem = as.character(mitem))%>%
     separate(c(1), c("mitem", "delete"))%>%
   select(!contains("delete"))%>%
    mutate(mitem =
             str_c(subject_item, mitem))

  }
  else if(subject == "math" && grade == 10){
```

```
    read_excel("_data/2022MCASDepartmentalAnalysis.xlsx", sheet = sheet_name,
             skip = 1, col_names= c(subject_item, "Type", "Reporting Category", "Standard
                  select(!contains("delete"))%>%
                  filter(!str_detect(mitem,"Legend|legend"))%>%
    mutate(mitem = as.character(mitem))%>%
    separate(c(1), c("mitem", "delete"))%>%
  select(!contains("delete"))%>%
   mutate(mitem =
            str_c(subject_item, mitem))

  }

}
```

```
## MCAS Preliminary Results Read In
read_MCAS_Prelim<-function(file_path, year){read_csv(file_path,
                   skip=1)%>%
  select(-c("sprp_dis", "sprp_sch", "sprp_dis_name", "sprp_sch_name", "sprp_orgtype",
  "schtype", "testschoolname", "yrsindis", "conenr_dis"))%>%

#Recode all nominal variables as characters


  mutate(testschoolcode = as.character(testschoolcode))%>%
  #Include this line when using the non-private dataframe
  # mutate(sasid = as.character(sasid))%>%
  mutate(highneeds = as.character(highneeds))%>%
  mutate(lowincome = as.character(lowincome))%>%
  mutate(title1 = as.character(title1))%>%
  mutate(ever_EL = as.character(ever_EL))%>%
  mutate(EL = as.character(EL))%>%
  mutate(EL_FormerEL = as.character(EL_FormerEL))%>%
  mutate(FormerEL = as.character(FormerEL))%>%
  mutate(ELfirstyear = as.character(ELfirstyear))%>%
  mutate(IEP = as.character(IEP))%>%
  mutate(plan504 = as.character(plan504))%>%
  mutate(firstlanguage = as.character(firstlanguage))%>%
  mutate(nature0fdis = as.character(natureofdis))%>%
  mutate(spedplacement = as.character(spedplacement))%>%
  mutate(town = as.character(town))%>%
  mutate(ssubject = as.character(ssubject))%>%
```

```
#Recode all ordinal variable as factors

  mutate(grade = as.factor(grade))%>%
  mutate(levelofneed = as.factor(levelofneed))%>%
  mutate(eperf2 = recode_factor(eperf2,
                                "E" = "Exceeding",
                                "M" = "Meeting",
                                "PM" = "Partially Meeting",
                                "NM"= "Not Meeting",
                               .ordered = TRUE))%>%
  mutate(eperflev = recode_factor(eperflev,
                                "E" = "E",
                                "M" = "M",
                                "PM" = "PM",
                                "NM"= "NM",
                                "DNT" = "DNT",
                                "ABS" = "ABS",
                               .ordered = TRUE))%>%
    mutate(mperf2 = recode_factor(mperf2,
                                "E" = "Exceeding",
                                "M" = "Meeting",
                                "PM" = "Partially Meeting",
                                "NM"= "Not Meeting",
                               .ordered = TRUE))%>%
  mutate(mperflev = recode_factor(mperflev,
                                "E" = "E",
                                "M" = "M",
                                "PM" = "PM",
                                "NM"= "NM",
                                "INV" = "INV",
                                "ABS" = "ABS",
                               .ordered = TRUE))%>%

  # The science variables contain a mixture of legacy performance levels and
  # next generation performance levels which needs to be addressed in the ordering
  # of these factors.
  mutate(sperf2 = recode_factor(sperflev,
                                "E" = "Exceeding",
                                "M" = "Meeting",
                                "PM" = "Partially Meeting",
```

```
                                    "NM"= "Not Meeting",
                                   .ordered = TRUE))%>%
  mutate(sperflev = recode_factor(sperf2,
                                    "E" = "E",
                                    "M" = "M",
                                    "PM" = "PM",
                                    "NM"= "NM",
                                    "INV" = "INV",
                                    "ABS" = "ABS",
                                   .ordered = TRUE))%>%
  #recode DOB using lubridate
  mutate(dob = mdy(dob,
  quiet = FALSE,
  tz = NULL,
  locale = Sys.getlocale("LC_TIME"),
  truncated = 0
))%>%
  mutate(IEP = case_when(
   IEP == "1" ~ "Disabled",
    IEP == "0" ~ "NonDisabled"
  ))%>%
    mutate(year = year)
}
```

```
## To-Do
##Function for number of items
## Function for graph for each question
## Function for key_word description
```

Data Read-In Tidy

Read in Student Performance Data

```
#Filter, rename variables, and mutate values of variables on read-in

MCAS_2022<-read_MCAS_Prelim("_data/PrivateSpring2022_MCAS_full_preliminary_results_0483030

#view(MCAS_2022)
head(MCAS_2022)
```

```
# A tibble: 6 x 256
  adminy~1 tests~2 grade grade~3 dob        gender race  yrsin~4 yrsin~5 yrsin~6
     <dbl> <chr>   <fct>   <dbl> <date>     <chr>  <chr> <chr>     <dbl>   <dbl>
1     2022 4830305 5           5 2010-10-04 M      W     3             3       1
2     2022 4830305 5           5 2010-09-24 F      W     5+            6       1
3     2022 4830305 5           5 2010-10-05 M      W     2             2       1
4     2022 4830305 5           5 2011-03-23 M      W     5+            6       1
5     2022 4830305 5           5 2011-10-17 F      W     2             2       1
6     2022 4830305 5           5 2011-04-30 M      W     5+            6       1
# ... with 246 more variables: highneeds <chr>, lowincome <chr>, title1 <chr>,
#   ever_EL <chr>, EL <chr>, EL_FormerEL <chr>, FormerEL <chr>,
#   ELfirstyear <chr>, IEP <chr>, plan504 <chr>, firstlanguage <chr>,
#   natureofdis <dbl>, levelofneed <fct>, spedplacement <chr>, town <chr>,
#   county <chr>, octenr <dbl>, conenr_sch <dbl>, conenr_sta <dbl>,
#   access_part <dbl>, ealt <lgl>, ecomplexity <lgl>, emode <chr>,
#   eteststat <chr>, wptopdev <lgl>, wpcompconv <lgl>, eitem1 <dbl>, ...
```

Introductory Physics, `SG9_Item` Read-In

```
# G9 Science Item analysis


SG9_Item<-read_item("SG9Physics", "science")%>%
  mutate(`Reporting Category` = case_when(
    `Reporting Category` == "EN" ~ "Energy",
    `Reporting Category` == "MF" ~ "Motion and Forces",
    `Reporting Category` == "WA" ~ "Waves"
    ))

head(SG9_Item)
```

```
# A tibble: 6 x 9
  sitem   Type  Reporting Cate~1 Stand~2 item ~3 item ~4 RT Pe~5 State~6 RT-St~7
  <chr>   <chr> <chr>            <chr>   <chr>     <dbl>   <dbl>   <dbl>   <dbl>
1 sitem20 CR    Energy           HS.PHY~ Interp~       4    0.45    0.53      -8
2 sitem37 CR    Energy           HS.PHY~ Descri~       3    0.48    0.52      -4
3 sitem6  SR    Energy           HS.PHY~ Descri~       2    0.51    0.41      10
4 sitem8  SR    Energy           HS.PHY~ Calcul~       1    0.43    0.52      -9
5 sitem12 SR    Energy           HS.PHY~ Descri~       1    0.66    0.68      -2
6 sitem14 SR    Energy           HS.PHY~ Calcul~       1    0.57    0.61      -4
# ... with abbreviated variable names 1: `Reporting Category`, 2: Standard,
#   3: `item Desc`, 4: `item Possible Points`, 5: `RT Percent Points`,
```

```
#   6: `State Percent Points`, 7: `RT-State Diff`
```

```
#view(SG9_Item)
```

Introductory Physics, `SG9_CU306Dis` Read-In

```
SG9_CU306Dis<-read_excel("_data/MCAS CU306 2022/CU306MCAS2022PhysicsGrade9ByDisability.xls
  sheet = "Disabled Students",
   col_names = c("Reporting Category", "Possible Points", "RT%Points",
                 "State%Points", "RT-State Diff"))%>%
  filter(`Reporting Category` == "Energy"|`Reporting Category`== "Motion, Forces, and Inte

#view(SG9_CU306Dis)


SG9_CU306Dis
```

```
# A tibble: 3 x 5
  `Reporting Category`             `Possible Points` `RT%Points` State~1 RT-St~2
  <chr>                            <chr>             <chr>       <chr>   <chr>
1 Energy                           18.0              0.38624     0.36627 2.0
2 Motion, Forces, and Interactions 30.0              0.43651     0.43807 0.0
3 Waves                            12.0              0.38095     0.33976 4.0
# ... with abbreviated variable names 1: `State%Points`, 2: `RT-State Diff`
```

Introductory Physics, `SG9_CU306NonDis` Read-In

```
SG9_CU306NonDis<-read_excel("_data/MCAS CU306 2022/CU306MCAS2022PhysicsGrade9ByDisability.
  sheet = "Non-Disabled Students",
   col_names = c("Reporting Category", "Possible Points", "RT%Points",
                 "State%Points", "RT-State Diff"))%>%
  filter(`Reporting Category` == "Energy"|`Reporting Category`== "Motion, Forces, and Inte

SG9_CU306NonDis
```

```
# A tibble: 3 x 5
  `Reporting Category`             `Possible Points` `RT%Points` State~1 RT-St~2
  <chr>                            <chr>             <chr>       <chr>   <chr>
1 Energy                           18.0              0.56501     0.59204 -2.0
2 Motion, Forces, and Interactions 30.0              0.60496     0.63597 -4.0
```

```
3 Waves                              12.0              0.4734      0.52223 -5.0
# ... with abbreviated variable names 1: `State%Points`, 2: `RT-State Diff`
```

```
#view(SG9_CU306NonDis)
```

Workflow Summary

After examining the summary of `MCAS_2022` (see appendix), I chose to

Filter:

- *SchoolID* : There are several variables that identify our school, I removed all but one, `testschoolcode`.
- *StudentPrivacy*: I left the `sasid` variable which is a student identifier number, but eliminated all values corresponding to students' names.
- `dis`: We are a charter school within our own unique district, therefore any "district level" data is identical to our "school level" data.

Rename

I currently have not renamed variables, but there are some trends to note:

- an `e` before most `ELA` MCAS student item performance metric variables
- an `m` before most `Math` MCAS student item performance metric variables
- an `s` before most `Science` MCAS student item performance metric variables

Mutate

I left as **doubles**

- variables that measured scores on specific MCAS items e.g., `mitem1`
- variables that measured student growth percentiles (`sgp`)
- variables that counted a student's years in the school system or state.

Recode to **char**

- variables that are **nominal** but have numeric values, e.g., `town`

Refactor as **ord**

- variables that are **ordinal**, e.g., `mperflev`.

Recode to **date**

- `dob` using lubridate.

Tidy Data and Join

I am interested in analyzing the 9th Grade Science Performance. To do this, I will select a subset of our `MCAS_2022` data frame which includes:

- 9th Grade students who took the Introductory Physics test
- Scores on the 42 Science Items
- points available on the
- Performance level on the test `sperflev`.
- Demographic characteristics of the students.

```
SG9_MCAS_2022 <- select(MCAS_2022, contains("sitem"), gender, grade, yrsinsch,
                                race, IEP, `plan504`, sattempt, sperflev, sperf2, sscaleds)%>
                               filter((grade == 9) & sattempt != "N")

SG9_MCAS_2022<-select(SG9_MCAS_2022, !(contains("43")|contains("44")|contains("45")))


#view(SG9_MCAS_2022)
head(SG9_MCAS_2022)
```

```
# A tibble: 6 x 52
  sitem1 sitem2 sitem3 sitem4 sitem5 sitem6 sitem7 sitem8 sitem9 sitem10 sitem11
   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>   <dbl>   <dbl>
1      1      0      0      0      1      1      1      0      1       0       0
2      1      1      1      1      1      1      0      1      0       1       1
3      1      0      0      0      1      2      1      1      1       0       0
4      1      1      1      0      1      0      1      0      1       0       0
5      1      1      1      1      1      2      1      1      0       0       1
6      1      1      0      0      0      1      1      0      0       0       1
# ... with 41 more variables: sitem12 <dbl>, sitem13 <dbl>, sitem14 <dbl>,
#   sitem15 <dbl>, sitem16 <dbl>, sitem17 <dbl>, sitem18 <dbl>, sitem19 <dbl>,
#   sitem20 <dbl>, sitem21 <dbl>, sitem22 <dbl>, sitem23 <dbl>, sitem24 <dbl>,
#   sitem25 <dbl>, sitem26 <dbl>, sitem27 <dbl>, sitem28 <dbl>, sitem29 <dbl>,
#   sitem30 <dbl>, sitem31 <dbl>, sitem32 <dbl>, sitem33 <dbl>, sitem34 <dbl>,
#   sitem35 <dbl>, sitem36 <dbl>, sitem37 <dbl>, sitem38 <dbl>, sitem39 <dbl>,
#   sitem40 <dbl>, sitem41 <dbl>, sitem42 <dbl>, gender <chr>, grade <fct>, ...
```

When I compared this data frame to the State reported analysis, the state analysis only contains 68 students. Notably, my data frame has 69 entries while the state is reporting data on only 68 students. I will have to investigate this further.

Since I will join this data frame with the `SG9_Item`, using `sitem` as the key, I need to pivot this data set longer.

```
SG9_MCAS_2022<- pivot_longer(SG9_MCAS_2022, contains("sitem"), names_to = "sitem", values_

#view(SG9_MCAS_2022)
head(SG9_MCAS_2022)
```

```
# A tibble: 6 x 12
  gender grade yrsinsch race  IEP   plan504 satte~1 sperf~2 sperf2 sscal~3 sitem
  <chr>  <fct>    <dbl> <chr> <chr> <chr>   <chr>   <ord>   <ord>    <dbl> <chr>
1 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
2 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
3 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
4 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
5 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
6 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
# ... with 1 more variable: sitem_score <dbl>, and abbreviated variable names
#   1: sattempt, 2: sperflev, 3: sscaleds
```

As expected, we now have 42 X 69 = 2898 rows.

Now, we should be ready to join our data sets using `sitem` as the key. We should have a 2,898 by (10 + 8) = 2,898 by 18 data frame. We will also check our raw data against the performance data reported by the state in the item report by calculating `percent_earned` by Rising Tide students and comparing it to the figure `RT Percent Points` and storing the difference in `earned_diff` ### Join and Sanity Checks

```
SG9_StudentItem <- SG9_MCAS_2022 %>%
  left_join(SG9_Item, "sitem")


head(SG9_StudentItem)
```

```
# A tibble: 6 x 20
  gender grade yrsinsch race  IEP   plan504 satte~1 sperf~2 sperf2 sscal~3 sitem
  <chr>  <fct>    <dbl> <chr> <chr> <chr>   <chr>   <ord>   <ord>    <dbl> <chr>
1 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
2 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
3 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
4 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
```

```
5 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
6 M      9            3 W     NonD~ 0       F       Partia~ Parti~     498 site~
# ... with 9 more variables: sitem_score <dbl>, Type <chr>,
#   `Reporting Category` <chr>, Standard <chr>, `item Desc` <chr>,
#   `item Possible Points` <dbl>, `RT Percent Points` <dbl>,
#   `State Percent Points` <dbl>, `RT-State Diff` <dbl>, and abbreviated
#   variable names 1: sattempt, 2: sperflev, 3: sscaleds
```

```
SG9_StudentItem
```

```
# A tibble: 2,898 x 20
   gender grade yrsin~1 race  IEP   plan504 satte~2 sperf~3 sperf2 sscal~4 sitem
   <chr>  <fct>   <dbl> <chr> <chr> <chr>   <chr>   <ord>   <ord>    <dbl> <chr>
 1 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 2 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 3 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 4 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 5 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 6 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 7 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 8 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
 9 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
10 M      9           3 W     NonD~ 0       F       Partia~ Parti~     498 site~
# ... with 2,888 more rows, 9 more variables: sitem_score <dbl>, Type <chr>,
#   `Reporting Category` <chr>, Standard <chr>, `item Desc` <chr>,
#   `item Possible Points` <dbl>, `RT Percent Points` <dbl>,
#   `State Percent Points` <dbl>, `RT-State Diff` <dbl>, and abbreviated
#   variable names 1: yrsinsch, 2: sattempt, 3: sperflev, 4: sscaleds
```

```
SG9_StudentItem%>%
   group_by(sitem)%>%
   summarise(percent_earned = round(sum(sitem_score, na.rm=TRUE)/sum(`item Possible Points
   left_join(SG9_Item, "sitem")%>%
   mutate(earned_diff = percent_earned-`RT Percent Points`)
```

```
# A tibble: 42 x 11
   sitem   perce~1 Type  Repor~2 Stand~3 item ~4 item ~5 RT Pe~6 State~7 RT-St~8
   <chr>     <dbl> <chr> <chr>   <chr>   <chr>     <dbl>   <dbl>   <dbl>   <dbl>
 1 sitem1     0.87 SR    Motion~ HS.PHY~ Analyz~       1    0.87    0.86       1
 2 sitem10    0.54 SR    Waves   HS.PHY~ Explai~       1    0.54    0.63      -9
```

```
 3 sitem11    0.46 SR    Motion~ HS.PHY~ Calcul~       1    0.47    0.65     -18
 4 sitem12    0.65 SR    Energy  HS.PHY~ Descri~       1    0.66    0.68      -2
 5 sitem13    0.84 SR    Motion~ HS.PHY~ Comple~       1    0.84    0.8        4
 6 sitem14    0.57 SR    Energy  HS.PHY~ Calcul~       1    0.57    0.61      -4
 7 sitem15    0.77 SR    Motion~ HS.PHY~ Identi~       2    0.78    0.71       7
 8 sitem16    0.49 CR    Motion~ HS.PHY~ Calcul~       3    0.5     0.46       4
 9 sitem17    0.7  SR    Motion~ HS.PHY~ Calcul~       1    0.71    0.71       0
10 sitem18    0.65 SR    Motion~ HS.PHY~ Descri~       1    0.66    0.66       0
# ... with 32 more rows, 1 more variable: earned_diff <dbl>, and abbreviated
#   variable names 1: percent_earned, 2: `Reporting Category`, 3: Standard,
#   4: `item Desc`, 5: `item Possible Points`, 6: `RT Percent Points`,
#   7: `State Percent Points`, 8: `RT-State Diff`
```

As expected, we now have a 2,898 X 18 data frame and the `earned_diff` values all round to 0.

G9 Science Performance Analysis

When considering our student performance data, we hope to address the following broad questions:

- What adjustments (if any) should be made at the Tier 1 level, i.e., curricular adjustments for all students?
- What would be the most beneficial areas of focus for a targeted intervention course for students struggling to meet or exceed performance expectations?
- Are there notable differences in student performance for students with and without disabilities?

Structure of the Exam

What reporting categories were emphasized by the state?

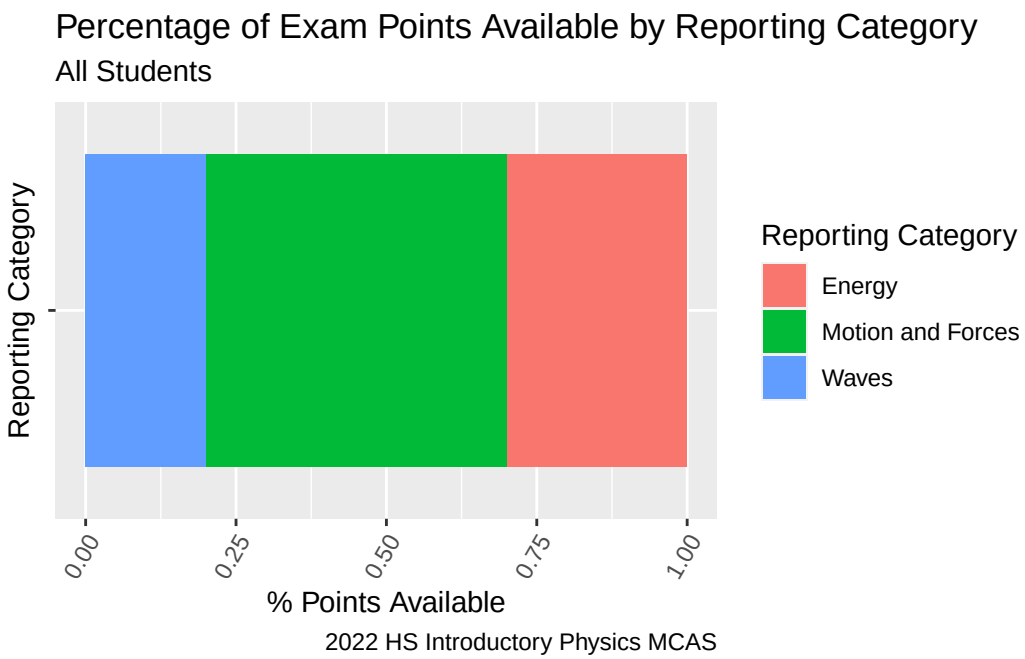
We can see from our summary that 50% of the exam points (30 of the available 60) come from questions from the Motion and Forces `Reporting Category`, followed by 30% from Energy, and 20% from Waves.

```
SG9_Cat_Total<-SG9_Item%>%
  select(`sitem`, `item Possible Points`, `Reporting Category`)%>%
  group_by(`Reporting Category`)%>%
  summarise(available_points = sum(`item Possible Points`, na.rm=TRUE))%>%
  mutate(percent_available_points = available_points/(sum(available_points, na.rm = TRUE))
```

```
SG9_Cat_Total
```

```
# A tibble: 3 x 3
  `Reporting Category` available_points percent_available_points
  <chr>                          <dbl>                     <dbl>
1 Energy                            18                       0.3
2 Motion and Forces                 30                       0.5
3 Waves                             12                       0.2
```

```
ggplot(SG9_Cat_Total, aes(x='',fill = `Reporting Category`, y = `available_points`)) +
    geom_bar(position="fill", stat = "identity") + coord_flip()+
 labs(subtitle ="All Students" ,
       y = "% Points Available",
       x= "Reporting Category",
       title = "Percentage of Exam Points Available by Reporting Category",
      caption = "2022 HS Introductory Physics MCAS")+
       theme(axis.text.x=element_text(angle=60,hjust=1))
```



Performance by Content Strands

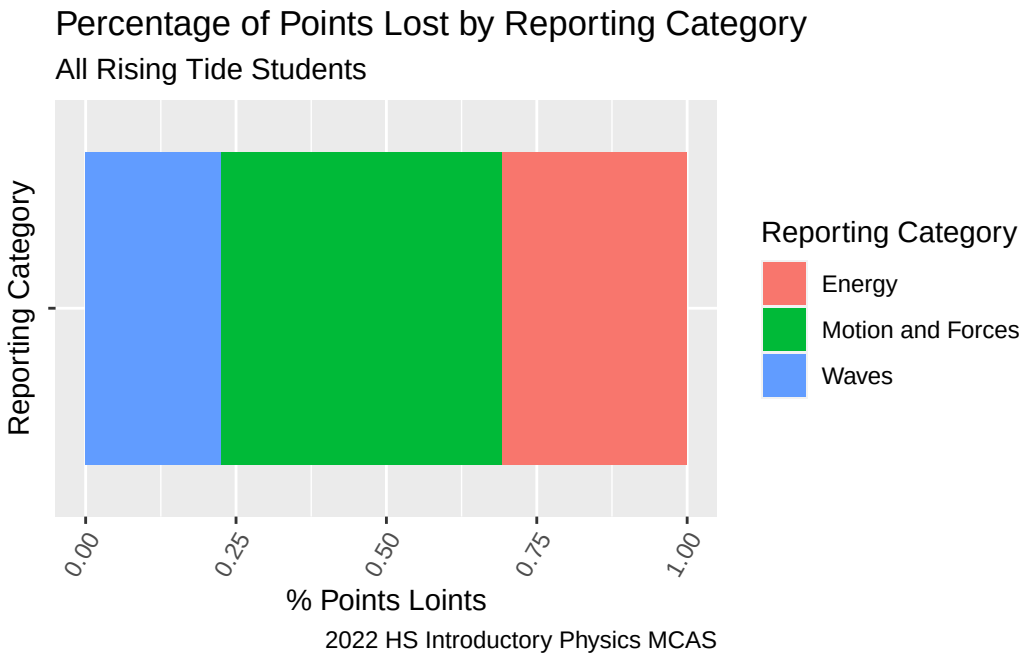
Where did RT students lose most of their points?

The proportion of points lost by Rising Tide students corresponds to the proportion of points available for each `Reporting Category` of the the exam. This suggests that our students are prepared consistently across the units in the `Reporting Categories`.

```
SG9_Cat_Loss<-SG9_StudentItem%>%
  select(`sitem`, `Reporting Category`, `item Possible Points`, `sitem_score`)%>%
  group_by(`Reporting Category`)%>%
  summarise(sum_points_lost = sum(`item Possible Points`-`sitem_score`, na.rm=TRUE),
            available_points = sum(`item Possible Points`, na.rm=TRUE))%>%
              mutate(percent_points_lost = round(sum_points_lost/sum(sum_points_lost,na.rm
            mutate(percent_available_points = available_points/(sum(available_points, na.r
SG9_Cat_Loss<-SG9_Cat_Loss%>%
  select(`Reporting Category`, `percent_available_points`, `percent_points_lost`)

SG9_Cat_Loss
```

```
# A tibble: 3 x 3
  `Reporting Category` percent_available_points percent_points_lost
  <chr>                                  <dbl>                <dbl>
1 Energy                                   0.3                 0.31
2 Motion and Forces                        0.5                 0.47
3 Waves                                    0.2                 0.22
```

```
SG9_Percent_Loss<-SG9_StudentItem%>%
  select(`sitem`, `Reporting Category`, `item Possible Points`, `sitem_score`)%>%
  mutate(`points_lost` = `item Possible Points` - `sitem_score`)%>%
  #ggplot(df, aes(x='', fill=option)) + geom_bar(position = "fill")
  ggplot( aes(x='',fill = `Reporting Category`, y = `points_lost`)) +
    geom_bar(position="fill", stat = "identity") + coord_flip()+
 labs(subtitle ="All Rising Tide Students" ,
       y = "% Points Loints",
       x= "Reporting Category",
       title = "Percentage of Points Lost by Reporting Category",
      caption = "2022 HS Introductory Physics MCAS")+
       theme(axis.text.x=element_text(angle=60,hjust=1))


SG9_Percent_Loss
```

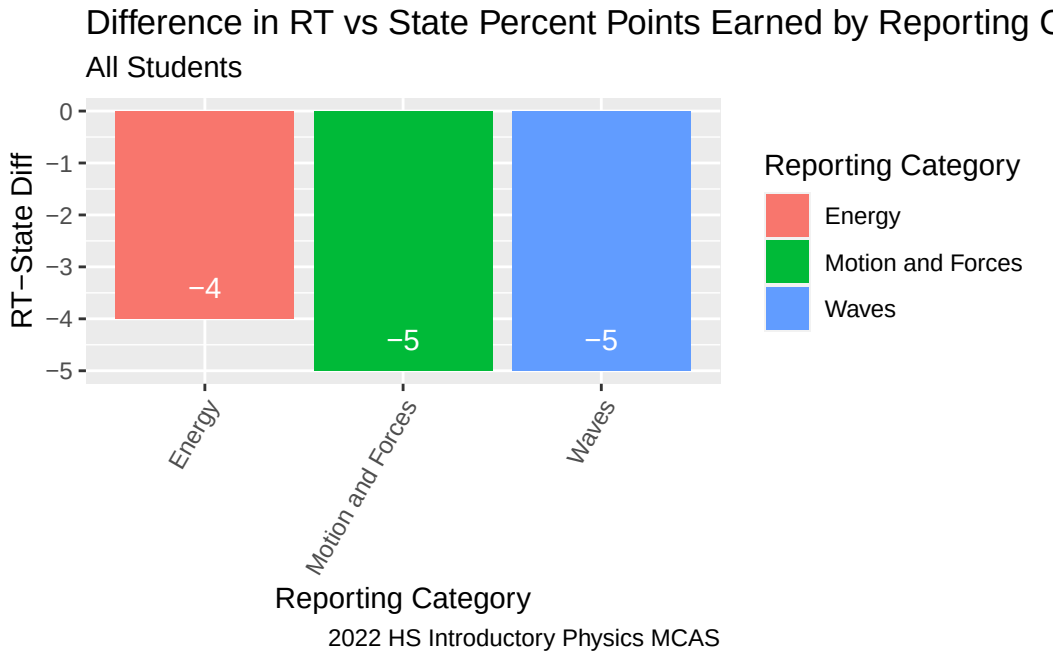


Did Rising Tide students' performance relative to the state vary by content reporting categories?

We can see from our table that on average our students earned between 4 and 5 percentage points fewer relative to their peers in the state on items across all three reporting categories.

```
SG9_Cat_RTState<-SG9_Item%>%
  select(`sitem`, `item Possible Points`, `Reporting Category`, `State Percent Points`, `R
  group_by(`Reporting Category`)%>%
  summarise(available_points = sum(`item Possible Points`, na.rm=TRUE),
            RT_points = sum(`RT Percent Points`*`item Possible Points`, na.rm = TRUE),
            RT_Percent_Points = 100*round(RT_points/available_points,2),
            State_Percent_Points = 100*round(sum(`State Percent Points`*`item Possible Poi
  mutate(`RT-State Diff` = round(RT_Percent_Points - State_Percent_Points, 2))%>%
 ggplot( aes(fill = `Reporting Category`, y=`RT-State Diff`, x=`Reporting Category`)) +
    geom_bar(position="dodge", stat="identity") +
  labs(subtitle ="All Students" ,
       y = "RT-State Diff",
       x= "Reporting Category",
       title = "Difference in RT vs State Percent Points Earned by Reporting Category",
      caption = "2022 HS Introductory Physics MCAS")+
       theme(axis.text.x=element_text(angle=60,hjust=1))+
```

```
   geom_text(aes(label = `RT-State Diff`), vjust = -1., colour = "white", position = posit

SG9_Cat_RTState
```




```
#SG9_CU306Dis
#SG9_CU306NonDis
```

```
# #view(SG9_Item)
# SG9_Cat_StateDiff<-SG9_Item%>%
#   select(`sitem`, `Reporting Category`, `item Possible Points`, `State Percent Points`,
#   group_by(`Reporting Category`)%>%
#   summarise(avg_RT_State_Diff = round(mean(`RT-State Diff`, na.rm=TRUE),2),
#             sd_RT_State_Diff = sd(`RT-State Diff`, na.rm=TRUE),
#             med_RT_State_Diff = median(`RT-State Diff`, na.rm=TRUE),
#             sum_RT_State_Diff = sum(`RT-State Diff`, na.rm=TRUE))
#
# SG9_Cat_StateDiff
```

Here we see the distribution of `RT-State Diff` (difference between the percentage of points earned on a given item by Rising Tide students and percentage of points earned on the same item by their peers in the State) by `sitem` and content `Reporting Category`. We can see

generally that items in the Motion and Forces `Reporting Category` seem to have the highest variability in student performance relative to the state. It would be worth looking at the specific question strands with the Physics Teachers. (It would be helpful to add item labels to the dots)

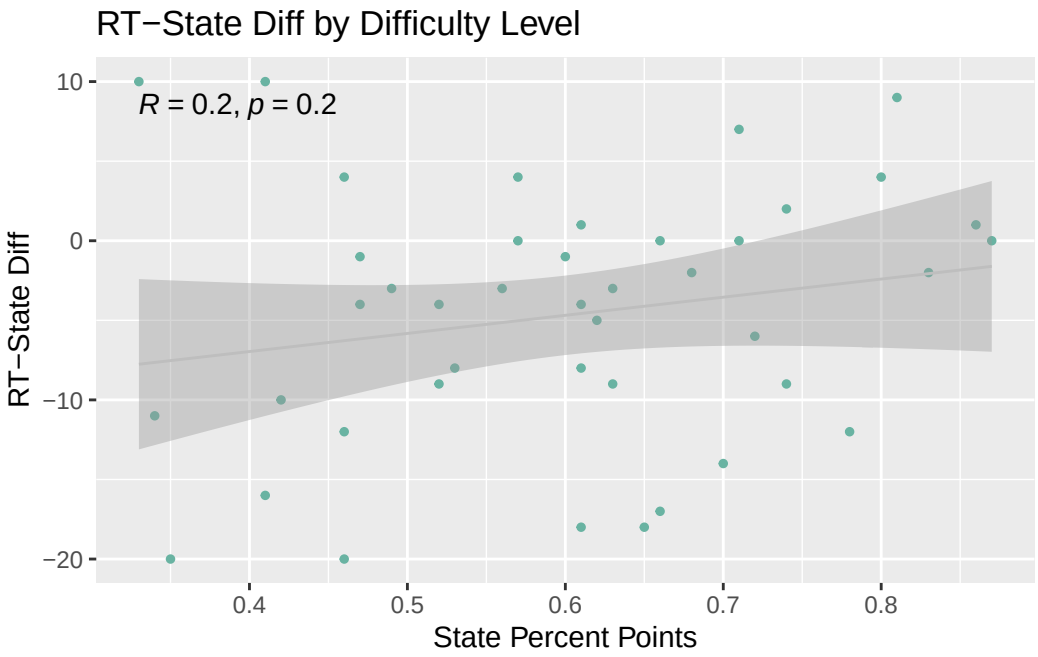
```
SG9_Cat_Box <-SG9_Item%>%
  select(`sitem`, `Reporting Category`, `State Percent Points`, `RT-State Diff`)%>%
  group_by(`Reporting Category`)%>%
  ggplot( aes(x=`Reporting Category`, y=`RT-State Diff`, fill=`Reporting Category`)) +
    geom_boxplot() +
    scale_fill_viridis(discrete = TRUE, alpha=0.6) +
    geom_jitter(color="black", size=0.1, alpha=0.9) +
    theme_ipsum() +
    theme(
      legend.position="none",
      plot.title = element_text(size=11)
    ) +
    ggtitle("G9 Introductory Physics School State Difference by Item") +
    xlab("")

ggplotly(SG9_Cat_Box)
```

Student Performance by Item Difficulty

Can differences in Rising Tide student performance on an item and State performance on an item be explained by the difficulty level of an item?

When considering `RT-State Diff` against `State Percent Points` for each `sitem` on the MCAS, this does not seem to generally be the case. Although the regression line shows `RT-State Diff` more likely to be negative on items where students in the State earned fewer points; the p-value is not significant.

```
G9Sci_Diff_Dot<-SG9_Item%>%
  select(`State Percent Points`, `RT-State Diff`, `Reporting Category`)%>%
 ggplot( aes(x=`State Percent Points`, y=`RT-State Diff`)) +
     geom_point(size = 1, color="#69b3a2")+
 geom_smooth(method="lm",color="grey", size =.5 )+
  labs(title = "RT-State Diff by Difficulty Level", y = "RT-State Diff",
       x = "State Percent Points") +
  stat_cor(method = "pearson")#+facet(vars(`Reporting Category`)) +#label.x = 450, label.y

G9Sci_Diff_Dot
```



Student Performance Key Words

How did students perform based on key words?

When scanning the `item Desc` entries in the `SG9_Item` data frame, there are several questions containing the word "Calculate" in their description.

How much is calculation emphasized on this exam and how did Rising Tide students perform relative to their peers in the state on items containing "calculate" in their description?

```
SG9_Calc<-SG9_Item%>%
  select(`sitem`, `item Desc`,`item Possible Points`, `Reporting Category`, `State Percent
   mutate( key_word = case_when(
     !str_detect(`item Desc`, "calculate|Calculate") ~ "Non-Calc",
     str_detect(`item Desc`, "calculate|Calculate") ~ "Calc"))
#view(SG9_Calc)

SG9_Calc
```

```
# A tibble: 42 x 7
   sitem   `item Desc`                    item ~1 Repor~2 State~3 RT-St~4 key_w~5
   <chr>   <chr>                            <dbl> <chr>      <dbl>   <dbl> <chr>
 1 sitem20 Interpret a diagram to deter~       4 Energy      0.53      -8 Calc
```

```
 2 sitem37 Describe how the average mol~       3 Energy     0.52      -4 Non-Ca~
 3 sitem6  Describe how the energy betw~       2 Energy     0.41      10 Non-Ca~
 4 sitem8  Calculate the efficiency of ~       1 Energy     0.52      -9 Calc
 5 sitem12 Describe how the energy of a~       1 Energy     0.68      -2 Non-Ca~
 6 sitem14 Calculate the kinetic energy~       1 Energy     0.61      -4 Calc
 7 sitem27 Describe how the mass and en~       1 Energy     0.81       9 Non-Ca~
 8 sitem33 Interpret a temperature vs. ~       1 Energy     0.87       0 Non-Ca~
 9 sitem34 Determine the observation th~       1 Energy     0.61     -18 Non-Ca~
10 sitem35 Calculate the final temperat~       1 Energy     0.35     -20 Calc
# ... with 32 more rows, and abbreviated variable names
#   1: `item Possible Points`, 2: `Reporting Category`,
#   3: `State Percent Points`, 4: `RT-State Diff`, 5: key_word
```

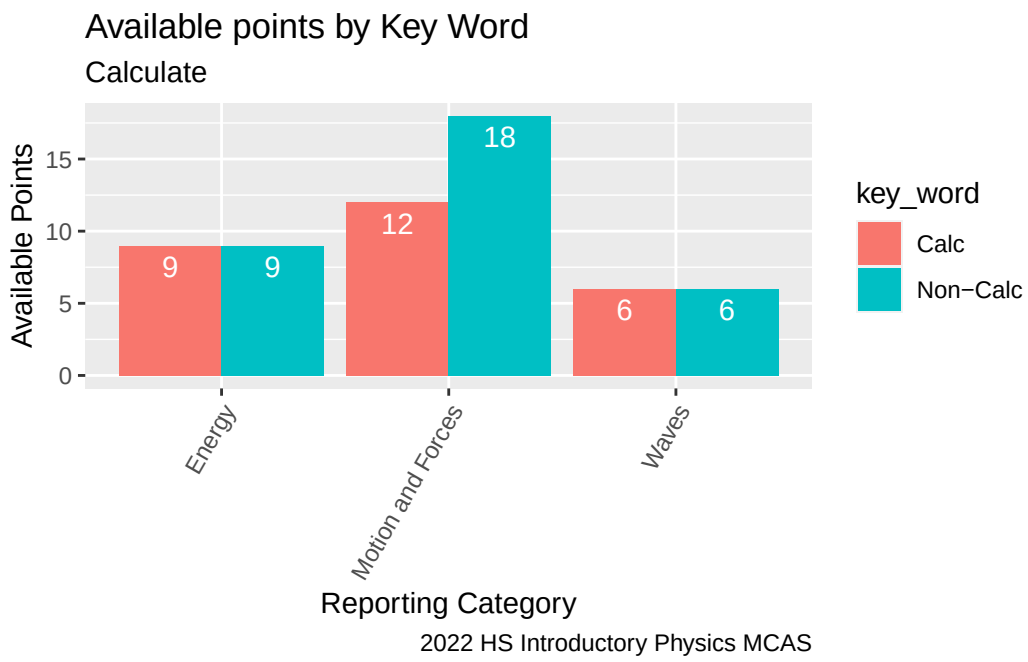
```
SG9_Calc%>%
  group_by(`Reporting Category`, `key_word`)%>%
  summarise(avg_RT_State_Diff = mean(`RT-State Diff`, na.rm=TRUE),
            med_RT_State_Diff = median(`RT-State Diff`, na.rm =TRUE),
            #sum_RT_State_Diff = sum(`RT-State Diff`, na.rm=TRUE),
            sum_sitem_Possible_Points = sum(`item Possible Points`, na.rm = TRUE))
```

```
# A tibble: 6 x 5
# Groups:   Reporting Category [3]
  `Reporting Category` key_word avg_RT_State_Diff med_RT_State_Diff sum_sitem_~1
  <chr>                <chr>                <dbl>             <dbl>        <dbl>
1 Energy               Calc                -8.8               -8              9
2 Energy               Non-Calc            -0.833             -1              9
3 Motion and Forces    Calc                -8                 -10            12
4 Motion and Forces    Non-Calc            -4.12               -1.5          18
5 Waves                Calc                -9.33              -9              6
6 Waves                Non-Calc             0.4                1              6
# ... with abbreviated variable name 1: sum_sitem_Possible_Points
```

Now, we can see that by the Waves and Energy categories half of the available points come from questions with calculate and half do not. In the Motion and Forces category, 40% of points are associated with questions that ask students to "calculate".

```
SG9_Calc_PointsAvail<-SG9_Calc%>%
  group_by(`Reporting Category`, `key_word`)%>%
  summarise(avg_RT_State_Diff = mean(`RT-State Diff`, na.rm=TRUE),
            med_RT_State_Diff = median(`RT-State Diff`, na.rm =TRUE),
            sum_RT_State_Diff = sum(`RT-State Diff`, na.rm=TRUE),
```

```
          sum_item_Possible_Points = sum(`item Possible Points`, na.rm = TRUE))%>%
   ggplot(aes(fill=`key_word`, y=sum_item_Possible_Points, x=`Reporting Category`)) + geom
  labs(subtitle ="Calculate" ,
       y = "Available Points",
       x= "Reporting Category",
       title = "Available points by Key Word",
      caption = "2022 HS Introductory Physics MCAS")+
       theme(axis.text.x=element_text(angle=60,hjust=1))+
   geom_text(aes(label = `sum_item_Possible_Points`), vjust = 1.5, colour = "white", posit

SG9_Calc_PointsAvail
```

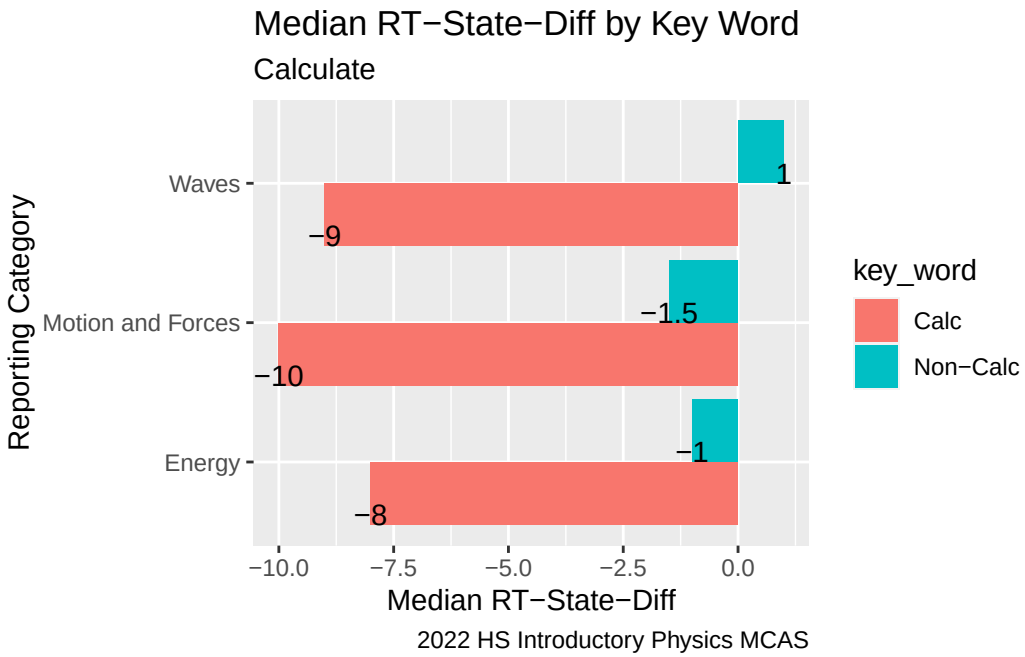



```
# SG9_Calc_PointsAvail_Stacked <- SG9_Calc%>%
#   ggplot(aes(fill=key_word, y = `item Possible Points`, x=`Reporting Category`)) +
#     geom_bar(position="stack", stat="identity")+
#   labs(subtitle = "Stacked Bar Chart",
#        y = "Available Points",
#       x= "Reporting Category",
#       title = "Available points by Key Word",
#      caption = "2022 HS Introductory Physics MCAS")+
#      theme(axis.text.x=element_text(angle=60,hjust=1))+
```

```
#   coord_flip()
#
# SG9_Calc_PointsAvail_Stacked
```

When we compare the median `RT-State Diff` for items containing the word "calculate" in their description vs. items that do not, we can see that across all of the `Reporting Categories` Rising Tide students performed significantly weaker relative to their peers in the state on questions that asked them to "calculate".

```
SG9_Calc_MedDiffBar<-SG9_Calc%>%
  group_by(`Reporting Category`, `key_word`)%>%
  summarise(mean_RT_State_Diff = round(mean(`RT-State Diff`, na.rm=TRUE),2),
            med_RT_State_Diff = median(`RT-State Diff`, na.rm =TRUE),
            sum_RT_State_Diff = sum(`RT-State Diff`, na.rm=TRUE))%>%
   ggplot(aes(fill=`key_word`, y=med_RT_State_Diff, x=`Reporting Category`)) + geom_bar(po
   labs(subtitle ="Calculate" ,
       y = "Median RT-State-Diff",
       x= "Reporting Category",
       title = "Median RT-State-Diff by Key Word",
      caption = "2022 HS Introductory Physics MCAS")+
   geom_text(aes(label = `med_RT_State_Diff`), vjust = 1.5, colour = "black", position = p
SG9_Calc_MedDiffBar
```



Here we can see the distribution of `RT-State Diff` by `sitem` and `Reporting Category` and the disparity in `RT-State Diff` when we consider items asking students to "Calculate" vs. those that do not.

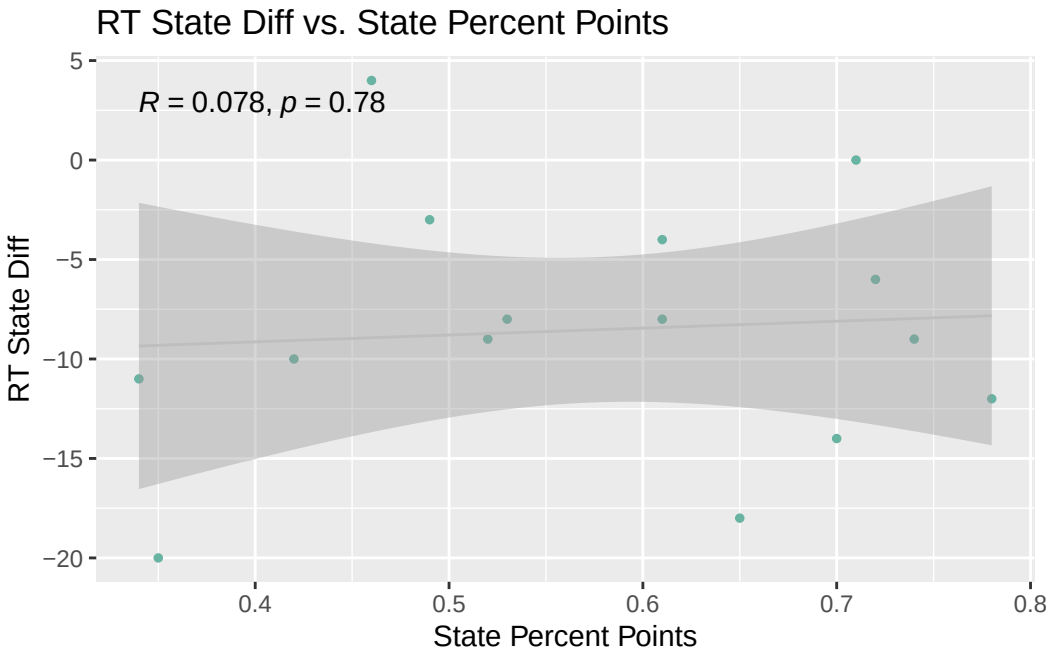
```
# SG9_Calc_Box <-SG9_Calc%>%
#   group_by(`key_word`, `Reporting Category`)%>%
#   ggplot( aes(x=`key_word`, y=`RT-State Diff`, fill=`Reporting Category`)) +
#     geom_boxplot() +
#     scale_fill_viridis(discrete = TRUE, alpha=0.6) +
#     geom_jitter(color="black", size=0.1, alpha=0.9) +
#     theme_ipsum() +
#     theme(
#       #legend.position="none",
#       plot.title = element_text(size=11)
#     ) + labs(subtitle ="Calculate" ,
#        y = "RT-State-Diff",
#        x= "Calculate vs. Non-Calculate",
#        title = "RT-State-Diff by Key Word",
#       caption = "2022 HS Introductory Physics MCAS")
#    # ggtitle("RT-State-Diff by Key Word") +
#   #  xlab("")
#
# SG9_Calc_Box
```

Did RT students perform worse relative to their peers in the state on more "challenging" calculation items?

If we consider the difficulty of items containing the word `calculate` for students as reflected in the state-wide performance (`State Percent Points`) for a given item, the gap between Rising Tide students' performance to their peers in the state `RT-State Diff` does not seem to increase significantly with the difficulty .

```
#view(SG9_Calc)
SG9_Calc_Dot<- SG9_Calc%>%
  select(`State Percent Points`, `RT-State Diff`, `key_word`)%>%
  filter(key_word == "Calc")%>%
  ggplot( aes(x=`State Percent Points`, y=`RT-State Diff`)) +
     geom_point(size = 1, color="#69b3a2")+
 geom_smooth(method="lm",color="grey", size =.5 )+
  labs(title = "RT State Diff vs. State Percent Points", y = "RT State Diff",
       x = "State Percent Points")+
    stat_cor(method = "pearson")
```

```
SG9_Calc_Dot
```



Is the "calculation gap" consistent across performance levels?

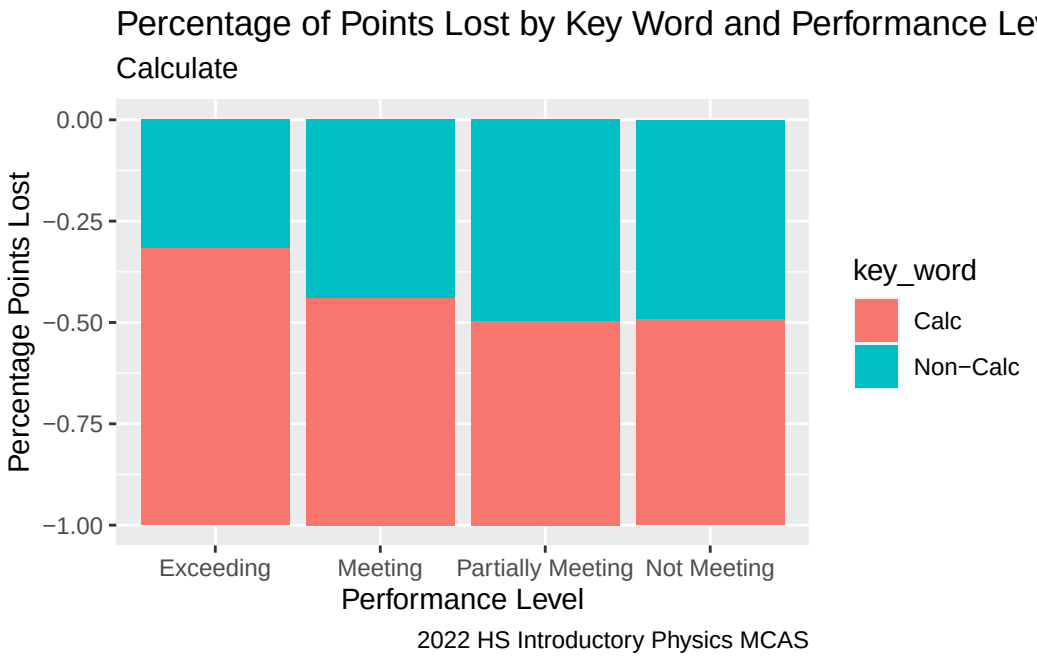
Here we can see that students with a higher performance level lost a greater proportion of their points on questions involving "Calculate". I.e., the higher a student's performance level, the greater the percentage of their points were lost to items asking them to "calculate". This suggests that in the general classroom to raise student performance, students should spend a higher proportion of time on calculation based activities.

```
# G9 Points Lost
G9Sci_StudentCalcPerflev<-SG9_StudentItem%>%
  select(gender, sitem, sitem_score, `item Desc`, `item Possible Points`, `State Percent P
  mutate( key_word = case_when(
     !str_detect(`item Desc`, "calculate|Calculate") ~ "Non-Calc",
     str_detect(`item Desc`, "calculate|Calculate") ~ "Calc"))%>%
  group_by(`sperflev`, `key_word`)%>%
  summarise(total_points_lost = sum(`sitem_score`-`item Possible Points`, na.rm = TRUE),
            med_RT_State_Diff = median(`RT-State Diff`, na.rm=TRUE))
G9Sci_StudentCalcPerflev
```

```
# A tibble: 8 x 4
# Groups:   sperflev [4]
  sperflev          key_word total_points_lost med_RT_State_Diff
  <ord>             <chr>                <dbl>             <dbl>
1 Exceeding         Calc                   -13                -9
2 Exceeding         Non-Calc                -6                -1
3 Meeting           Calc                  -232                -9
4 Meeting           Non-Calc              -182                -1
5 Partially Meeting Calc                  -611                -9
6 Partially Meeting Non-Calc              -602                -1
7 Not Meeting       Calc                  -171                -9
8 Not Meeting       Non-Calc              -164                -1
```

```
#view(SG9_StudentItem)


# G9Sci_StudentCalc<-SG9_StudentItem%>%
#   select(gender, sitem, sitem_score, `item Desc`, `item Possible Points`, `State Percent
#   mutate( key_word = case_when(
#      !str_detect(`item Desc`, "calculate|Calculate") ~ "Non-Calc",
#      str_detect(`item Desc`, "calculate|Calculate") ~ "Calc"))%>%
#   group_by(`Reporting Category`, `key_word`)%>%
#   summarise(total_points_lost = sum(`sitem_score`-`item Possible Points`, na.rm = TRUE))
#  ggplot(aes(fill=`key_word`, y=total_points_lost, x=`Reporting Category`)) + geom_bar(po
#    labs(subtitle ="Calculate" ,
#        y = "Sum Points Lost",
#        x= "Reporting Category",
#        title = "Sum Points Lost by Key Word",
#       caption = "2022 HS Introductory Physics MCAS")+
#    geom_text(aes(label = `total_points_lost`), vjust = 1.5, colour = "black", position =


#+
   #geom_text(aes(label = `total_points_lost`), vjust = 1.5, colour = "white", position =

#G9Sci_StudentCalc




G9Sci_StudentCalcPerflev%>%
```

```
ggplot(aes(fill=`key_word`, y=total_points_lost, x=`sperflev`)) + geom_bar(position="fill
  labs(subtitle ="Calculate" ,
       y = "Percentage Points Lost",
       x= "Performance Level",
       title = "Percentage of Points Lost by Key Word and Performance Level",
      caption = "2022 HS Introductory Physics MCAS")
```




```
G9Sci_StudentCalcPerflev
```

```
# A tibble: 8 x 4
# Groups:   sperflev [4]
  sperflev          key_word total_points_lost med_RT_State_Diff
  <ord>             <chr>                <dbl>              <dbl>
1 Exceeding         Calc                   -13                 -9
2 Exceeding         Non-Calc                -6                 -1
3 Meeting           Calc                  -232                 -9
4 Meeting           Non-Calc              -182                 -1
5 Partially Meeting Calc                  -611                 -9
6 Partially Meeting Non-Calc              -602                 -1
7 Not Meeting       Calc                  -171                 -9
8 Not Meeting       Non-Calc              -164                 -1
```

Student Performance and Disability

Did RT students perform worse relative to their peers in the state on more "challenging" calculation items?

We can see from our `CU306` reports that our students with disabilities performed better relative to their peers in the state, `RT-State Diff`, across all `Reporting Categories`, while our non-disabled students performed worse relative to their peers in the state across all `Reporting Categories`. This suggest that more attention needs to be paid to the needs of the non-disabled students in the General Education setting.

```
SG9_CU306Dis%>%
  select(`RT-State Diff`, `Reporting Category`)%>%
  mutate(`Disability Satus` = "Disabled")
```

```
# A tibble: 3 x 3
  `RT-State Diff` `Reporting Category`             `Disability Satus`
  <chr>           <chr>                            <chr>
1 2.0             Energy                           Disabled
2 0.0             Motion, Forces, and Interactions Disabled
3 4.0             Waves                            Disabled
```

```
SG9_CU306NonDis%>%
  select(`RT-State Diff`, `Reporting Category`)%>%
  mutate(`Disability Satus` = "Non-Disabled")
```

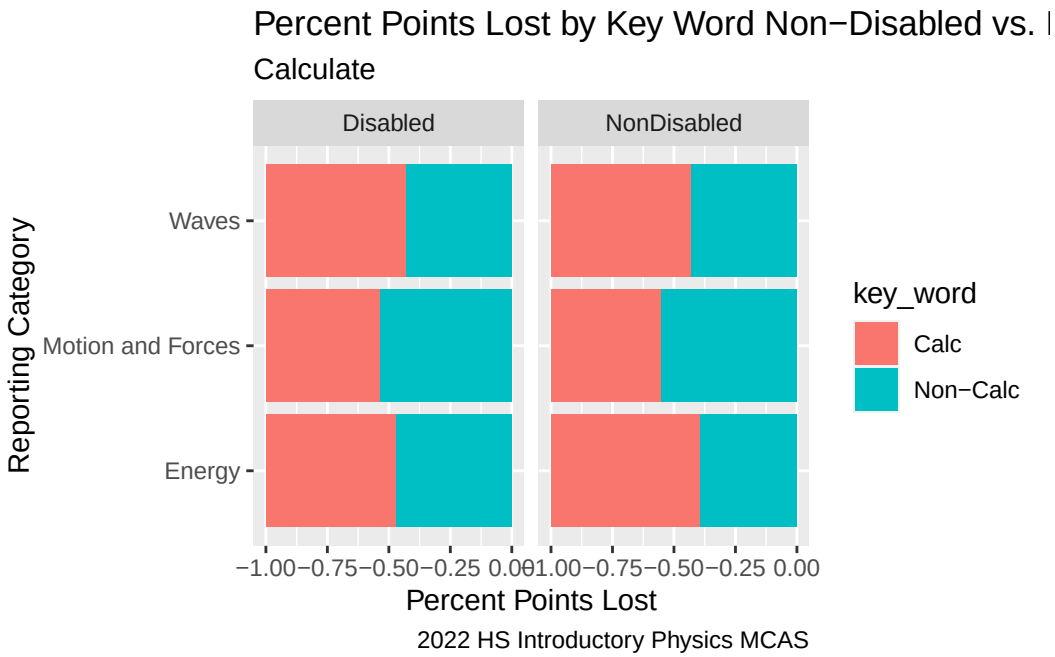
```
# A tibble: 3 x 3
  `RT-State Diff` `Reporting Category`             `Disability Satus`
  <chr>           <chr>                            <chr>
1 -2.0            Energy                           Non-Disabled
2 -4.0            Motion, Forces, and Interactions Non-Disabled
3 -5.0            Waves                            Non-Disabled
```

When we examine the points lost by reporting category and disability status, there does not seem to be a significant difference in performance between disabled and non-disabled students across `Reporting Categories`.

```
G9Sci_StudentCalcDis<-SG9_StudentItem%>%
  select(gender, sitem, sitem_score, `item Desc`, `item Possible Points`, `State Percent P
  mutate( key_word = case_when(
     !str_detect(`item Desc`, "calculate|Calculate") ~ "Non-Calc",
```

```
    str_detect(`item Desc`, "calculate|Calculate") ~ "Calc"))%>%
  group_by(`Reporting Category`, `key_word`, `IEP`)%>%
  summarise(total_points_lost = sum(`sitem_score`-`item Possible Points`, na.rm = TRUE))%>
 ggplot(aes(fill=`key_word`, y=total_points_lost, x=`Reporting Category`)) + geom_bar(posi
  facet_wrap(vars(IEP))+ coord_flip()+
   labs(subtitle ="Calculate" ,
       y = "Sum Points Lost",
       x= "Reporting Category",
       title = "Sum Points Lost by Key Word Non-Disabled vs. Disabled",
      caption = "2022 HS Introductory Physics MCAS")+
  geom_text(aes(label = `total_points_lost`), vjust = 1.5, colour = "black", position = po

#G9Sci_StudentCalcDis
```

```
G9Sci_StudentCalcDis<-SG9_StudentItem%>%
  select(gender, sitem, sitem_score, `item Desc`, `item Possible Points`, `State Percent P
  mutate( key_word = case_when(
     !str_detect(`item Desc`, "calculate|Calculate") ~ "Non-Calc",
     str_detect(`item Desc`, "calculate|Calculate") ~ "Calc"))%>%
  group_by(`Reporting Category`, `key_word`, `IEP`)%>%
  summarise(sum_points_lost = sum(`sitem_score`-`item Possible Points`, na.rm = TRUE))%>%
 ggplot(aes(fill=`key_word`, y=sum_points_lost, x=`Reporting Category`)) + geom_bar(positi
  facet_wrap(vars(IEP))+ coord_flip()+
   labs(subtitle ="Calculate" ,
       y = "Percent Points Lost",
       x= "Reporting Category",
       title = "Percent Points Lost by Key Word Non-Disabled vs. Disabled Students",
      caption = "2022 HS Introductory Physics MCAS")

G9Sci_StudentCalcDis
```



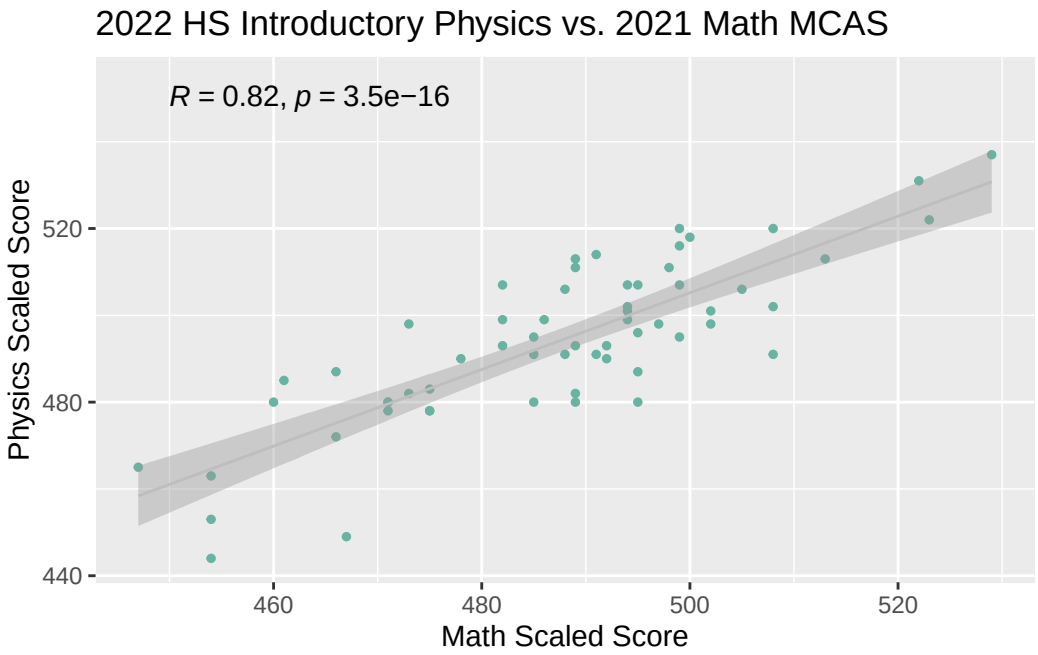

```
# view(SG9_CU306Dis)
# SG9_CU306Dis%>%
#   mutate(Disability_Status = "Disabled")
# SG9_CU306NonDis%>%
#   mutate(Disability_Status = "NonDisabled")
# SG9_CU306Dis<-SG9_CU306Dis%>%
#   full_join(SG9_CU306NonDis, "RT-State Diff")
# view(SG9_CU306Dis)
```

Conclusion

A student's performance on their 9th Grade Introductory Physics MCAS is strongly associated with their performance on their 8th Grade Math MCAS exam. This suggests, that the use of prior Math MCAS and current STAR Math testing data can identify students in need of extra support.

```
SG9_Math<-MCAS_2022%>%
  select(sscaleds, mscaleds2021,sscaleds_prior, grade, sattempt)%>%
  filter((grade == 9) & sattempt != "N")%>%
  ggplot(aes(x=`mscaleds2021`, y =`sscaleds`))+
```

```
    geom_point(size = 1, color="#69b3a2")+
 geom_smooth(method="lm",color="grey", size =.5 )+
  labs(title = "2022 HS Introductory Physics vs. 2021 Math MCAS", y = "Physics Scaled Scor
       x = "Math Scaled Score") +
  stat_cor(method = "pearson", label.x = 450, label.y = 550)

SG9_Math
```



Rising Tide students as a whole performed slightly weaker relative to the state in all content reporting areas; however, disabled students performed better relative to their peers in the state. The performance gap is accounted for by the performance of the non-disabled students in the general classroom setting.

All Rising Tide students, regardless of disability status, performed significantly weaker relative to students in the State on items including the key word Calculate in their `item description`. This suggests, we should dedicate more classroom instructional time to problem solving with calculation. Notably, the higher a student's performance level, the higher the percentage of points a student lost for calculation items. To increase the proportion of students exceeding expectations, we need to improve our students' performance on calculation based items; evidence based math interventions include small group, differentiated problem sets.

The discrepancy in performance by Rising Tide students with and without disabilities further supports the need for differentiated small group problem sets in the general classroom setting

and tiered, small-group, problem sets for non-disabled students.

Reflection: Limitations/Areas for Improvement

-Reflection (2-3 pages) Describe your process/experience with the project, including the decisions you made, what was most challenging, and what you wish you would have known. You can also discuss what the next steps would be, were you to continue with the project.

To improve this report, I need to

- edit down the number of visualizations used.
- edit the average points lost by category to get average over all items, rather than average of averages.
- mutate `IEP` so that the values `0` and `1` are `Disabled` and `Non-Disabled`
- mutate `Reporting Category` so that the full category names appear in the graphs.
- be more discerning on when to use `totals` and when to use `averages`. To improve performance on a test, we are concerned with total points lost; to identify curricular weaknesses we are also interested in relative performance to the state by content area

References

Chang, W. (2022). *R Graphics Cookbook, 2nd Edition*. O'Reilly Media.

Grolemund, G., & Wickham, H. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media.

HighSchool Introductory Physics Item Report [Data] https://profiles.doe.mass.edu/mcas/mcasitems2.aspx?grade=HS&subjectcode=PHY&linkid=23&orgcode=04830000&fycode=2022&orgtypecode=5&

H. Wickham. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2009.

Papay, J. P., Mantil, A., McDonough, A., Donahue, K., An, L., & Murnane, R. J. (n.d.). ***Lifting all boats? Accomplishments and Challenges from 20 Years of Education Reform in Massachusetts***. Retrieved December 2, 2022, from https://annenberg.brown.edu/sites/default/files/LiftingAllBoats_FINAL.pdf

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.https://www.r-project.org.

RStudio Team. (2019). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA. https://www.rstudio.com.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.

Appendix

Codebook MCAS_2022 Variables

DESE

variable	Measurement Level	Values
`gender`	Nominal	the reported gender identify of the student. Female: F, Male: M, Non-binary: N
`item Description`	Nominal	details of assessment question
`item Possible Points`	Discrete	The number of points available for a given `sitem`
`Reporting Category`	Nominal	content area of `sitem` Motion and Forces Waves Energy
`RT Percent Points`	Continuous	Percent of points earned by Rising Tide Students for a given `sitem`
`RT-State Diff`	Discrete	Difference between percent of points earned by Rising Tide Students and Students in the State for a given `sitem`
`sitem`	Nominal	The question number the MCAS exam
`sitem_score`	Discrete	The number of points a student earned on a given `sitem`
`sperflev`	Ordinal	Exceeds Expectations Meets Expectations Partially Meets Expectations Does Not Meet Expectations
`sscaleds`	Discrete	The student's scaled score by subject area (e: English, m: Math, s: Science)
`ssgp`	Continuous	The student's growth percentile by subject area (e: English, m: Math, s: Science)
`State Percent Points`	Continuous	Percent of points earned by Massachusetts students for a given `sitem`

MCAS 2022 Data Summary

```
# examine the summary to decide how to best set up our data frame

 print(summarytools::dfSummary(MCAS_2022,
```