Pipe Dreams Are Made of These

Dependency

```
library(tidyverse)
```

Read CSV file

```
students <-read_csv("./Resources/students.csv")
```

```
## Parsed with column specification:
## cols(
##   `Student ID` = col_double(),
##   student_name = col_character(),
##   gender = col_character(),
##   grade = col_character(),
##   school_name = col_character(),
##   reading_score = col_double(),
##   math_score = col_double()
## )
```

```
schools <- read_csv("./Resources/schools.csv")
```

```
## Parsed with column specification:
## cols(
##   `School ID` = col_double(),
##   school_name = col_character(),
##   type = col_character(),
##   size = col_double(),
##   budget = col_double()
## )
```

Preview tibble

```
students %>% head()
```

```
## # A tibble: 6 x 7
##   `Student ID` student_name   gender grade school_name  reading_score math_score
##          <dbl> <chr>          <chr>  <chr> <chr>                <dbl>      <dbl>
## 1            0 Paul Bradley   M      9th   Huang High ~            66         79
## 2            1 Victor Smith   M      12th  Huang High ~            94         61
```

```
## 3           2 Kevin Rodrigu~ M      12th  Huang High ~          90         60
## 4           3 Dr. Richard S~ M      12th  Huang High ~          67         58
## 5           4 Bonnie Ray     F      9th   Huang High ~          97         84
## 6           5 Bryan Miranda  M      9th   Huang High ~          94         94
```

```
schools %>% head()
```

```
## # A tibble: 6 x 5
##   `School ID` school_name           type      size  budget
##         <dbl> <chr>                 <chr>    <dbl>   <dbl>
## 1           0 Huang High School     District  2917 1910635
## 2           1 Figueroa High School  District  2949 1884411
## 3           2 Shelton High School   Charter   1761 1056600
## 4           3 Hernandez High School District  4635 3022020
## 5           4 Griffin High School   Charter   1468  917500
## 6           5 Wilson High School    Charter   2283 1319574
```

Join two tibbles

```
data2 = left_join(students, schools, by=c("school_name"))
data2 %>% head()
```

```
## # A tibble: 6 x 11
##   `Student ID` student_name gender grade school_name reading_score math_score
##          <dbl> <chr>        <chr>  <chr> <chr>                <dbl>      <dbl>
## 1           0 Paul Bradley M      9th   Huang High~             66         79
## 2           1 Victor Smith M      12th  Huang High~             94         61
## 3           2 Kevin Rodri~ M      12th  Huang High~             90         60
## 4           3 Dr. Richard~ M      12th  Huang High~             67         58
## 5           4 Bonnie Ray   F      9th   Huang High~             97         84
## 6           5 Bryan Miran~ M      9th   Huang High~             94         94
## # ... with 4 more variables: `School ID` <dbl>, type <chr>, size <dbl>,
## #   budget <dbl>
```

Total Number of Schools

```
school_count <- students$school_name %>%
  unique() %>%
  length()

school_count
```

```
## [1] 15
```

Total Number of Students

```
student_count <-  students %>% nrow()
student_count
```

```
## [1] 39170
```

Average reading and math scores

```
mean_reading_score <- summarize(students, mean(reading_score))
mean_math_score <- summarize(students, mean(math_score))
```

Calculate the percentage of students with passing reading scores, i.e. over 70%.

```
percentage_passing_reading <- students %>%
  filter(reading_score > 70) %>%
  nrow() * 100 / student_count %>%
  round(2)

percentage_passing_reading
```

```
## [1] 82.97166
```

Calculate the percentage of students with passing math scores, i.e. over 70%.

```
percentage_passing_math <-  students %>%
  filter(math_score > 70) %>%
  nrow() * 100 / student_count %>%
  round(2)

percentage_passing_math
```

```
## [1] 72.39214
```

Calculate the overall passing rate, i.e. the average of math and reading passing percentages

```
overall_passing_rate <- (percentage_passing_math + percentage_passing_reading) / 2
overall_passing_rate
```

```
## [1] 77.6819
```

Calculate the average math and reading scores by school

```
students %>%
  group_by(school_name) %>%
  summarize(avg.reading=mean(reading_score), avg.math=mean(math_score))
```

```
## # A tibble: 15 x 3
##    school_name           avg.reading avg.math
##    <chr>                       <dbl>    <dbl>
```

```
##  1 Bailey High School           81.0     77.0
##  2 Cabrera High School          84.0     83.1
##  3 Figueroa High School         81.2     76.7
##  4 Ford High School             80.7     77.1
##  5 Griffin High School          83.8     83.4
##  6 Hernandez High School        80.9     77.3
##  7 Holden High School           83.8     83.8
##  8 Huang High School            81.2     76.6
##  9 Johnson High School          81.0     77.1
## 10 Pena High School             84.0     83.8
## 11 Rodriguez High School        80.7     76.8
## 12 Shelton High School          83.7     83.4
## 13 Thomas High School           83.8     83.4
## 14 Wilson High School           84.0     83.3
## 15 Wright High School           84.0     83.7
```

Calculate the average math and reading scores by grade level at each school

```
students %>%
  group_by(school_name, grade) %>%
  summarize(avg.reading=mean(reading_score), avg.math=mean(math_score))
```

```
## # A tibble: 60 x 4
## # Groups:   school_name [15]
##    school_name          grade avg.reading avg.math
##    <chr>                <chr>       <dbl>    <dbl>
##  1 Bailey High School   10th         80.9     77.0
##  2 Bailey High School   11th         80.9     77.5
##  3 Bailey High School   12th         80.9     76.5
##  4 Bailey High School   9th          81.3     77.1
##  5 Cabrera High School  10th         84.3     83.2
##  6 Cabrera High School  11th         83.8     82.8
##  7 Cabrera High School  12th         84.3     83.3
##  8 Cabrera High School  9th          83.7     83.1
##  9 Figueroa High School 10th         81.4     76.5
## 10 Figueroa High School 11th         80.6     76.9
## # ... with 50 more rows
```

```
total_budget <- schools %>%
  summarize(sum(budget))
```

Display data

```
paste("School count: ", school_count)
```

```
## [1] "School count:  15"
```

```r
paste("Student count: ", student_count)
```

```
## [1] "Student count:  39170"
```

```r
paste("Total budget: ", total_budget)
```

```
## [1] "Total budget:  24649428"
```

```r
paste("Average reading score: ", mean_reading_score)
```

```
## [1] "Average reading score:  81.8778401838141"
```

```r
paste("Average math score: ", mean_math_score)
```

```
## [1] "Average math score:  78.9853714577483"
```

```r
paste("% passing reading: ", percentage_passing_reading)
```

```
## [1] "% passing reading:  82.9716619862139"
```

```r
paste("% passing math: ", percentage_passing_math)
```

```
## [1] "% passing math:  72.3921368394179"
```

```r
paste("Overall passing rate: ", overall_passing_rate)
```

```
## [1] "Overall passing rate:  77.6818994128159"
```

Use sapply() to convert data type

```r
total_budget <- total_budget %>% sapply(as.numeric)
mean_math_score <- mean_math_score %>% sapply(as.numeric)
mean_reading_score <- mean_reading_score %>% sapply(as.numeric)
```

Create a per-school summary

```r
#part 1
school_summary_tb <- data2 %>%
  group_by(type, school_name) %>%
  summarise(Mean_Reading_Score = mean(reading_score),
            Mean_Math_Score    = mean(math_score),
            Total_Students=n(),
            Budget = mean(budget),
            Per_Student_Budget = mean(budget) / n()
            )

school_summary_tb %>% head(15)
```

```
## # A tibble: 15 x 7
## # Groups:   type [2]
##    type  school_name Mean_Reading_Sc~ Mean_Math_Score Total_Students Budget
##    <chr> <chr>                  <dbl>           <dbl>          <int>  <dbl>
##  1 Char~ Cabrera Hi~             84.0            83.1           1858 1.08e6
##  2 Char~ Griffin Hi~             83.8            83.4           1468 9.18e5
##  3 Char~ Holden Hig~             83.8            83.8            427 2.48e5
##  4 Char~ Pena High ~             84.0            83.8            962 5.86e5
##  5 Char~ Shelton Hi~             83.7            83.4           1761 1.06e6
##  6 Char~ Thomas Hig~             83.8            83.4           1635 1.04e6
##  7 Char~ Wilson Hig~             84.0            83.3           2283 1.32e6
##  8 Char~ Wright Hig~             84.0            83.7           1800 1.05e6
##  9 Dist~ Bailey Hig~             81.0            77.0           4976 3.12e6
## 10 Dist~ Figueroa H~             81.2            76.7           2949 1.88e6
## 11 Dist~ Ford High ~             80.7            77.1           2739 1.76e6
## 12 Dist~ Hernandez ~             80.9            77.3           4635 3.02e6
## 13 Dist~ Huang High~             81.2            76.6           2917 1.91e6
## 14 Dist~ Johnson Hi~             81.0            77.1           4761 3.09e6
## 15 Dist~ Rodriguez ~             80.7            76.8           3999 2.55e6
## # ... with 1 more variable: Per_Student_Budget <dbl>
```

Create Tibble of District Summary

```
#part 2
district_summary <- tribble(
  ~Total_Schools,
  ~Total_Students,
  ~Total_Budget,
  ~Avg_Math,
  ~Avg_Reading,
  ~Math_Passing_Pct,
  ~Reading_Passing_Pct,
  ~Total_Passing_Pct,
  school_count, student_count, total_budget, mean_reading_score, mean_math_score, percentage_passing_rea
)
```

Display summary of district-wide data

```
district_summary
```

```
## # A tibble: 1 x 8
##   Total_Schools Total_Students Total_Budget Avg_Math Avg_Reading
##           <int>          <int>        <dbl>    <dbl>       <dbl>
## 1            15          39170     24649428     81.9        79.0
## # ... with 3 more variables: Math_Passing_Pct <dbl>, Reading_Passing_Pct <dbl>,
## #   Total_Passing_Pct <dbl>
```