Table manipulation with dplyr

Load data and package

```
# load dplyr
library("dplyr")

# load information about students and teachers
student = read.csv("student.csv")
teacher = read.csv("teacher.csv")
```

Filter Rows

```
# Filter individual row
filter(student, teacherID == "T1" & year2score < 80)

# Filter groups
student %>% group_by(teacherID) %>% filter(mean(year1score) > 65)
```

Select columns

```
# Select columns
select(student, studentID, birthDay)

# Exclude columns
select(student, -studentID, -birthDay)
```

Sort rows

```
# Small to large
arrange(student, year1score)

# Large to small
arrange(student, desc(year1score))
```

Group rows

```
# Grouping
student %>% group_by(teacherID)

# Ungrouping
student %>% ungroup()
```

Summarise by group

```r
student %>% group_by(teacherID) %>%
  summarise(mean_score1 = mean(year1score),
            mean_score2 = mean(year2score))
```

Add or edit column

```r
# Add new column
student %>%
  mutate(mean_score = (year1score + year2score)/2)

# Edit existing column
student %>%
  mutate(year1score = scale(year1score))
```

Identify Unique rows

```r
# unique values of all columns
student %>% distinct()

# unique values of teacherID
student %>% distinct(teacherID)
```

Sample Rows

```r
# sample 10 rows
sample_n(student, 10)

# sample 30% of the rows
sample_frac(student, 0.3)
```

Load data and package

```
# load dplyr
library("dplyr")

# load information about students and teachers
student = read.csv("student.csv")
teacher = read.csv("teacher.csv")
```

Sample Rows

```
# sample 10 rows
sample_n(student, 10)

# sample 30% of the rows
sample_frac(student, 0.3)
```