hw 7

Cory Costello

I'm interested in how grades have changed (or stayed the same) across the state. I'm also interested in looking at this for different sections (or subjects; math, writing, etc.) and different measurements (i.e., growth vs. achievement)

Additionally, I'm going to work at the district-level. The main reason I'm doing this is because it seems like one of the highest levels to work with (as opposed to schools), and I don't want to deal with the dependencies at lower levels (e.g., dependencies between schools in a district). I imagine that even districts have dependendencies that someone more well-versed in education work would model (e.g., adjacency? or county?), but it seems more reasonable to treat districts as independent.

Moreover, I want to see if average grades and change in grades are the same in districts with more or less people on free and reduced lunch.

Finally, I'm going to limit my analyses to traditional public schools, and will eliminate charter and online schools before I aggregate to district-level data.

Load Data

First, I'll load the data. I'm going to load the final grades data from each year (2010, 2011, 2012) and the free and reduced lunch data from each year (2010, 2011, 2012).

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(rio)
library(ggthemes)

final_grades_2010 <- import("../../../Data/CO_data/2010_final_grade.csv",
                            setclass = "tbl_df") %>%
  janitor::clean_names()

final_grades_2011 <- import("../../../Data/CO_data/2011_final_grade.csv",
                            setclass = "tbl_df") %>%
  janitor::clean_names()

final_grades_2012 <- import("../../../Data/CO_data/2012_final_grade.csv",
                            setclass = "tbl_df") %>%
  janitor::clean_names()

# Free and reduced lunch data
```

```r
frl_2010 <- import("../../../Data/CO_data/2010_k_12_FRL.csv",
                              setclass = "tbl_df") %>%
  janitor::clean_names()

frl_2011 <- import("../../../Data/CO_data/2011_k_12_FRL.csv",
                              setclass = "tbl_df") %>%
  janitor::clean_names()

frl_2012 <- import("../../../Data/CO_data/2012_k_12_FRL.csv",
                              setclass = "tbl_df") %>%
  janitor::clean_names()
```

Prepare to filter out charter and online schools

As stated above, I'm going to limit my analyses to traditional public schools, so I'll eliminate charter and online schools. To do so, I'm going to use `anti_join()` to do a *filter join*.

To prepare for this, I'm going to set up a dataframe of just the charter and online schools from 2010 (the only year in which that data appears to be available). My assumption is that charter/online status is invariant across this period, though admittedly its possible that this will exclude some schools that shouldn't be excluded (if they were an online school / had an online program that went away in 2011)

```r
charter_and_online_schools_2010 <- final_grades_2010 %>%
  select(schoolname, schoolnumber, charteronline) %>%
  filter(charteronline == "Charter" |
           charteronline == "Charter & Online" |
           charteronline == "Online")
```

Tidy, Aggregate, and join Data

I'm going to tidy and aggregate the final grades datasets from each year in the same pipeline. This will essentially boil down to:

- filter join using `anti_join()` to remove charter and online schools
- Selecting just the following columns:
  - districtname - name of the school district
  - anything that contains growth - this will select all of the growth grade columns
  - anything that contains "ach" - this will select all of the achievement grade columns
- Renaming two columns:
  - `overall_weighted_growth_grade` will be renamed `overall_growth_grade` to match the other variables' patterns (for gathering)
  - `districtname` will be renamed `district_name`
    - This is already the name of that variable in `final_grades_2012`, so I'm changing the 2010 and 2011 datasets to match the 2012 one.
- gather the growth and achievement grades
  - I'm doing this by gathering everything except district_name (leaving only the growth and achievement grades).
- separate the newly gathered `variable` column into its three parts, which are:
  - `section` - this is the subject area of the grade (math, writing, science, etc.)
  - `type` - this is growth or achievement
  - `elim` - this repeats grade at each row; it will be removed

- selecting out the `elim` column
- grouping by district name, section (or subject), and type (achievement vs. growth)
- summarizing to get the mean (across schools within each district) grade for each section and measurement type (achievement vs. growth).
- mutate to add the year for the dataset (which will be helful once the three are joined)

```
district_lvl_finalgrades2010 <- final_grades_2010 %>%
  anti_join(charter_and_online_schools_2010) %>%
  select(districtname, contains("growth"), contains("ach")) %>%
  rename(overall_growth_grade = overall_weighted_growth_grade,
         district_name = districtname) %>%
  gather(variable, grade, -district_name) %>%
  separate(variable, c("section", "type", "elim")) %>%
  select(-elim) %>%
  group_by(district_name, section, type) %>%
  summarize(mean_grade = mean(grade, na.rm = TRUE)) %>%
  mutate(year = 2010)
```

```
## Joining, by = c("schoolname", "schoolnumber", "charteroronline")
```

```
district_lvl_finalgrades2011 <- final_grades_2011 %>%
  anti_join(charter_and_online_schools_2010) %>%
  select(districtname, contains("growth"), contains("ach")) %>%
  rename(overall_growth_grade = overall_weighted_growth_grade,
         district_name = districtname) %>%
  gather(variable, grade, -district_name) %>%
  separate(variable, c("section", "type", "elim")) %>%
  select(-elim) %>%
  group_by(district_name, section, type) %>%
  summarize(mean_grade = mean(grade, na.rm = TRUE)) %>%
  mutate(year = 2011)
```

```
## Joining, by = c("schoolname", "schoolnumber")
```

```
district_lvl_finalgrades2012 <- final_grades_2012 %>%
  # school name appears as school_name here, and schoolname elsewhere
  # (including the charteronline data used in anti_join)
  # going to rename it in this one, since its being thrown out anyway
  rename(schoolname = school_name) %>%
  anti_join(charter_and_online_schools_2010) %>%
  select(district_name, contains("growth"), contains("ach")) %>%
  rename(overall_growth_grade = overall_weighted_growth_grade) %>%
  gather(variable, grade, -district_name) %>%
  separate(variable, c("section", "type", "elim")) %>%
  select(-elim) %>%
  group_by(district_name, section, type) %>%
  summarize(mean_grade = mean(grade, na.rm = TRUE)) %>%
  mutate(year = 2012)
```

```
## Joining, by = "schoolname"
```

Next, I'm going to use rbind to combine the three tidy district-level final grades data.

```
district_lvl_finalgrades_allyears <- rbind(district_lvl_finalgrades2010, district_lvl_finalgrades2011)
  rbind(district_lvl_finalgrades2012)
```

Now I'll tidy and aggregate (to district level) the percentage of students that receive free and reduced lunch.

This boils down to:

- changing `percent_free_reduced` variable from character to double (with `parse_number()` and `mutate()`).
- rename schoolname to school_name
- filter out charter and online schools with anti_join
- group by district name
- aggregate to district level data with summarize, taking the average percentage of students on free and reduced lunch per district.
- Mutate to add year
    - Also going to mutate to bin free and reduced lunch into the following categories:
        - grand mean
        - · 1 SD
        - · 1 SD

```
district_lvl_frl_2010 <- frl_2010 %>%
  mutate(percent_free_and_reduced = parse_number(percent_free_and_reduced)) %>%
  rename(schoolname = school_name) %>%
  anti_join(charter_and_online_schools_2010) %>%
  group_by(district_name) %>%
  summarize(m_pct_frl = mean(percent_free_and_reduced, na.rm = TRUE)) %>%
  mutate(year = 2010,
         frl_bin = ifelse(m_pct_frl > (mean(m_pct_frl, na.rm = TRUE)
                                       + sd(m_pct_frl, na.rm = TRUE)), "high",
                          ifelse(m_pct_frl < (mean(m_pct_frl, na.rm = TRUE)
                                       - sd(m_pct_frl, na.rm = TRUE)), "low",
                          "medium")))
```

```
## Joining, by = "schoolname"
```

```
district_lvl_frl_2011 <- frl_2011 %>%
  mutate(percent_free_and_reduced = parse_number(percent_free_and_reduced)) %>%
  rename(schoolname = school_name) %>%
  anti_join(charter_and_online_schools_2010) %>%
  group_by(district_name) %>%
  summarize(m_pct_frl = mean(percent_free_and_reduced, na.rm = TRUE)) %>%
  mutate(year = 2011,
         frl_bin = ifelse(m_pct_frl > (mean(m_pct_frl, na.rm = TRUE)
                                       + sd(m_pct_frl, na.rm = TRUE)), "high",
                          ifelse(m_pct_frl < (mean(m_pct_frl, na.rm = TRUE)
                                       - sd(m_pct_frl, na.rm = TRUE)), "low",
                          "medium")))
```

```
## Joining, by = "schoolname"
```

```
district_lvl_frl_2012 <- frl_2012 %>%
  mutate(percent_free_and_reduced = parse_number(percent_free_and_reduced)) %>%
  rename(schoolname = school_name) %>%
  anti_join(charter_and_online_schools_2010) %>%
  group_by(district_name) %>%
  summarize(m_pct_frl = mean(percent_free_and_reduced, na.rm = TRUE)) %>%
  mutate(year = 2012,
         frl_bin = ifelse(m_pct_frl > (mean(m_pct_frl, na.rm = TRUE)
                                       + sd(m_pct_frl, na.rm = TRUE)), "high",
                          ifelse(m_pct_frl < (mean(m_pct_frl, na.rm = TRUE)
                                       - sd(m_pct_frl, na.rm = TRUE)), "low",
                          "medium")))
```

```
## Joining, by = "schoolname"
```

Now I'll join these free and reduced lunch data to each other. I'm also going to relevel the percent free lunch bins to make sure they're low = 1, medium = 2, and high = 3

```r
district_lvl_frl_allyears <- rbind(district_lvl_frl_2010,
                                   district_lvl_frl_2011) %>%
  rbind(district_lvl_frl_2012) %>%
  mutate(frl_bin = factor(frl_bin, levels = c("low", "medium", "high")))
```

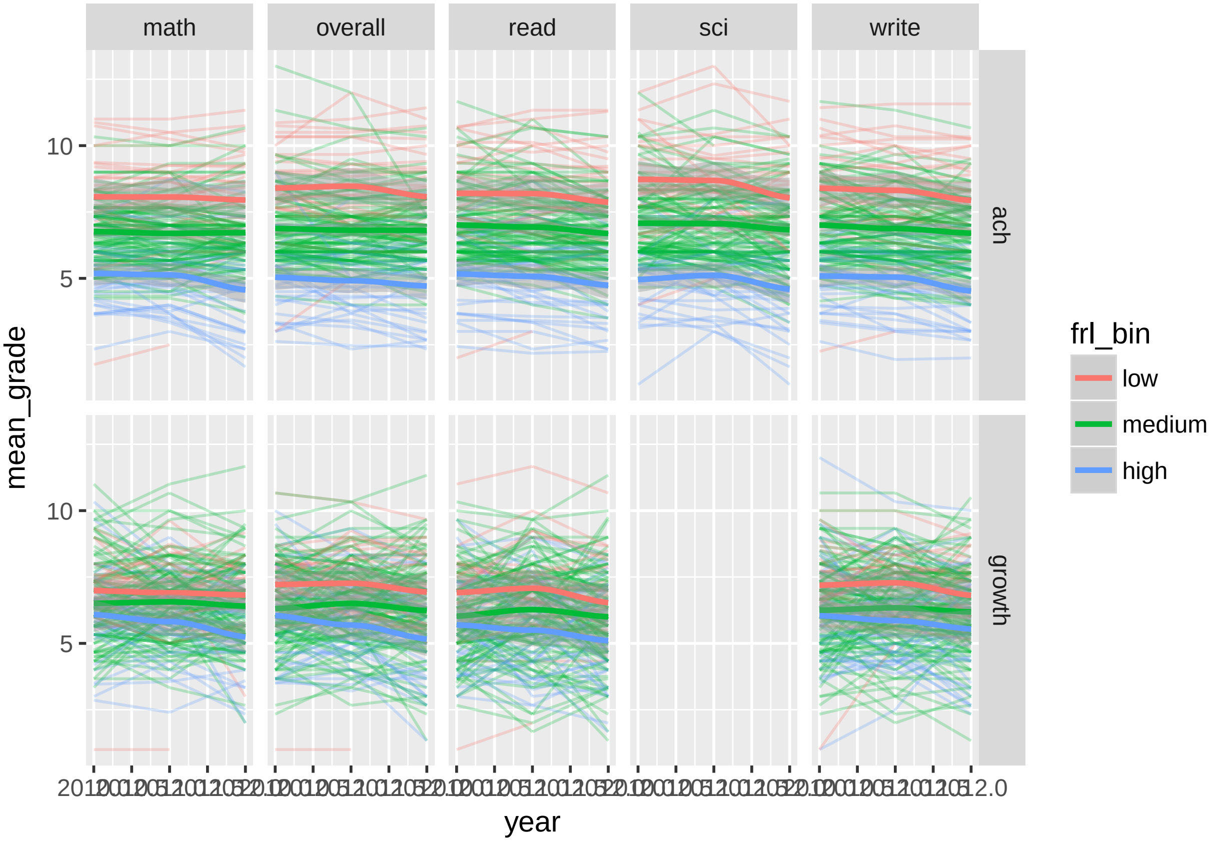
And finally, join the district level grades data and free and reduced lunch data.

```r
district_lvl_grades_frl_allyears <- district_lvl_finalgrades_allyears %>%
  left_join(district_lvl_frl_allyears) %>%
  ungroup()
```

```
## Joining, by = c("district_name", "year")
```

Plot 1: Changes in Grades across time by subject, measurement type, and percentage of students that receive free and reduced lunch

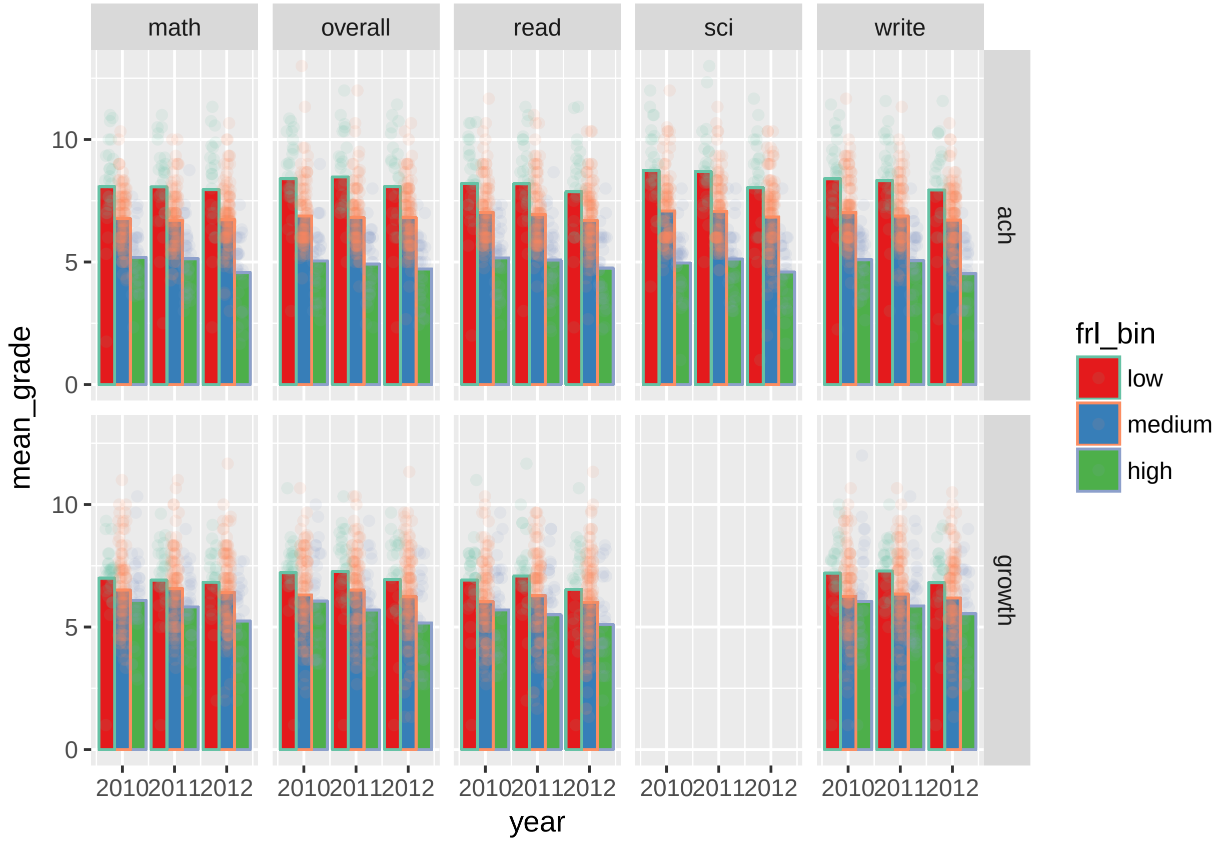
```r
district_lvl_grades_frl_allyears %>%
  filter(!is.na(frl_bin) &
           section != "ell") %>%
ggplot(aes(x = year, y = mean_grade, color = frl_bin)) +
  #geom_point() +
  geom_line(aes(group = district_name), alpha = .25) +
  geom_smooth( size = 1) +
  facet_grid(type ~ section)
```



Plot 2: Average grade by year, subject, and measurement type

```
district_lvl_grades_frl_allyears %>%
  filter(!is.na(frl_bin) &
           section != "ell") %>%
  group_by(frl_bin, section, type, year) %>%
  mutate(frlbin_mean_grade = mean(mean_grade, na.rm = TRUE)) %>%

  ggplot(aes(x = year, y = mean_grade, color = frl_bin)) +
  geom_bar(aes(y = frlbin_mean_grade, fill = frl_bin),
           stat = "identity", position = "dodge")+
  geom_point(position = position_jitterdodge(), alpha = .1) +
  facet_grid(type ~ section) +
  scale_fill_brewer(palette = "Set1") +
  scale_color_brewer(palette = "Set2")
```



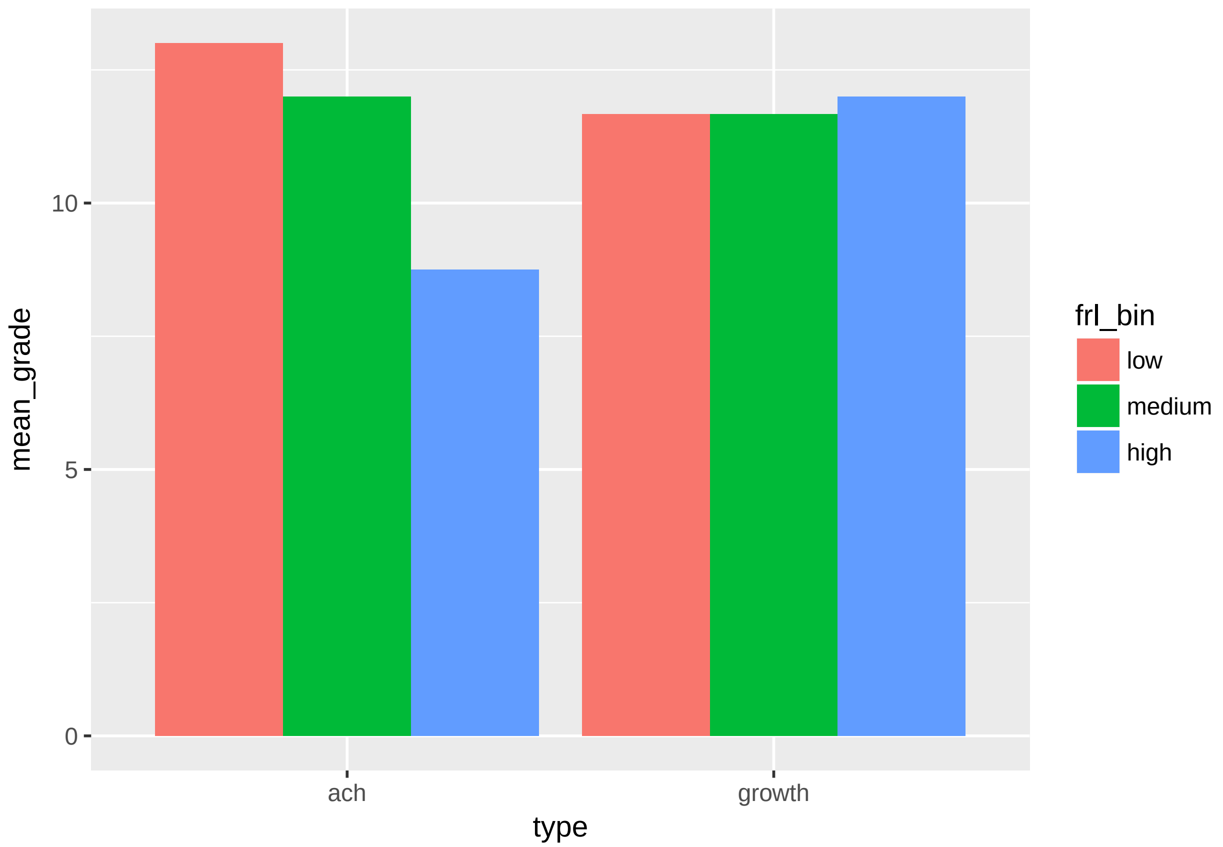
It looks like differences in grade by percentage of students in free and reduced lunch are less strong with the growth factor than the achievement factor. I'm going to attempt to see that a little clearer in the following plot.

Sicne I'll be collapsing across section/subject, I'll remove the overall scores (don't want those to go into the mean across sections).

Plot 3: Average grade by percentage of students on free and reduced lunch (high, medium, low) for achievement and growth

```
district_lvl_grades_frl_allyears %>%
  filter(!is.na(frl_bin) &
           section != "ell" &
           section != "overall") %>%
  ungroup() %>%
  ggplot(aes(x = type, y = mean_grade, fill = frl_bin)) +
  geom_bar(stat = "identity", position = "dodge")
```

```
## Warning: Removed 47 rows containing missing values (geom_bar).
```



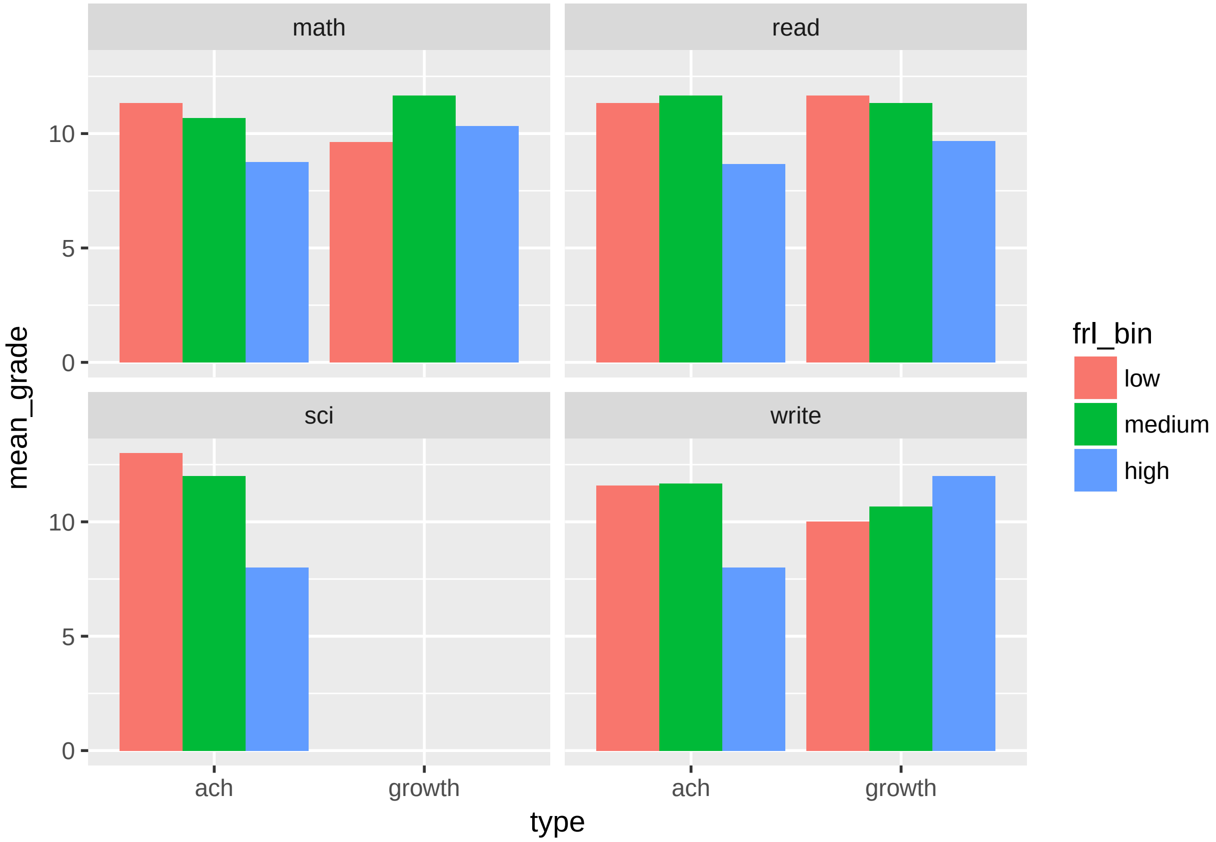
Interesting, it looks like there basically now differences in growth based on percentage of students in free and reduced lunch, and it looks like achievement is only worse for districts with a high percentage of students on free and reduced lunch (low and medium look the same)

Next, let's see this by subject

Plot 4: Average Grade by Measurement Type and Percentage of Students on Free and Reduced Lunch in District

```
district_lvl_grades_frl_allyears %>%
  filter(!is.na(frl_bin) &
           section != "ell" &
           section != "overall") %>%
  ungroup() %>%
  ggplot(aes(x = type, y = mean_grade, fill = frl_bin)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~section)
```

```
## Warning: Removed 47 rows containing missing values (geom_bar).
```

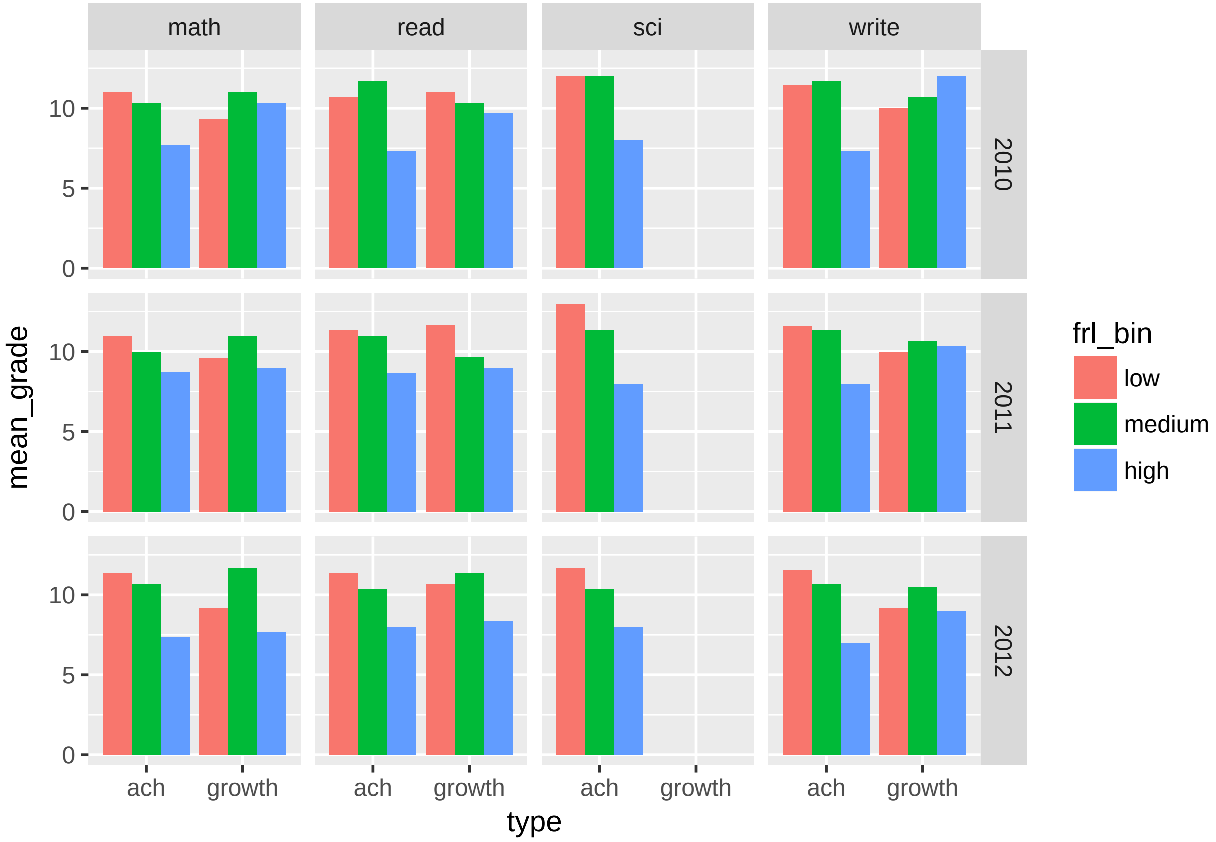


Hmm, looks like districts with a high percentage of students on free and reduced lunch perform consistently worse when assessed by achievement. When assessed by growth, they perform similarly on math, a little lower on reading, and a little better on writing. Now let's see if this is similar in different years.

Plot 5: Average Grade by Measurement Type and Percentage of Students on Free and Reduced Lunch in District by Year

```
district_lvl_grades_frl_allyears %>%
  filter(!is.na(frl_bin) &
           section != "ell" &
           section != "overall") %>%
  ungroup() %>%
  ggplot(aes(x = type, y = mean_grade, fill = frl_bin)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_grid(year~section)
```

```
## Warning: Removed 47 rows containing missing values (geom_bar).
```



Looks like the pattern is consistent across years for achievement (where grades in districts with a high percentage of students on free and reduced lunch perform worse), but is much less consistent across years for growth.