Data Team Learning R Module 2 Exercise: Reproducible Reporting with ggplot2 and RMarkdown

September 2016

Context[1]

In the past few years, The Metro City School District (MCSD) has lost teachers to neighboring districts and charter school networks. One of the primary reasons teachers cited for leaving MCSD, especially among early-career and high-performing teachers, was the existing compensation structure.

MCSD uses a "steps-and-lanes" compensation structure typical of public school districts nationwide. In this system, teachers' compensation is based on experience ("steps") and highest degree attained ("lanes").

MCSD wants to know if their current compensation system rewards its highest performing teachers and has contracted with TNTP to evaluate its teacher compensation system.

Data

We have provided salary and teacher background information from MCSD teachers from the 2015-2016 school year. The data is a .csv file. This data contains the following variables:

teacher_id: Numeric

experience: Years of teaching experience, including the current year (i.e., first year teachers have a value of 1)

exp_cat: MCSD's classification of teachers into groups by experience. 1 are New teachers (1st and 2nd years), 2 are Novice teachers (3rd – 6th year), 3 are Mid-Career (7th – 15th year), 4 are Experienced teachers (16th – 24th year), and 5 are Career teachers (25th year and beyond).

degree: Highest degree earned.

salary: Numeric, in dollars.

score: Final evaluation score, on a scale of 100 to 300.

rating: Final evaluation categorical rating. Ratings are based solely on one's score. Scores less than 150 are "Ineffective", 150 – 199.9 are "Minimally Effective", 200 – 249.9 are "Developing", 251 – 274.9 are "Proficient", and 275 –300 are "Highly Effective"

retention: Voluntary retention. 1 = the teacher returned for the following year, 0 = they left the district.

[1] This exercise is loosely based on an exercise TNTP has used to screen candidates for compensation modeling analysts, though the data is different.

Research Questions

Your goal is to write a brief memo for TNTP's project director using RMarkdown answering MCSD's initial research questions. The answers to question 0 are provided as a working example.

0. *[Example] What's the distribution of ratings in MCSD?*
    a. Run the provided code – it will create a table when you knit the RMarkdown document.
    b. Code is also provided to show this distribution in a chart.

1. *Confirm the problem*. Is the district losing its highest-performing teachers?
    a. Create a table showing the retention rate by final evaluation rating.
    b. Write 1-2 sentences summarizing the state of the district's differential retention.[2]

2. *Explore the relationship between teacher experience and performance*.
    a. Create a figure showing the relationship between years of experience and evaluation score.
    b. Write 1-2 sentences summarizing this relationship.

3. *Document the (lack of) relationship between highest degree held and teacher performance*.
    a. Create a figure showing the relationship between degree and evaluation score.
    b. Write 1-2 sentences summarizing this relationship.

4. *Explore the connection between performance and compensation*.
    a. Create a figure showing the relationship between evaluation score and salary.
    b. Write 1-2 sentences summarizing the relationship between performance and compensation.

5. *BONUS: Contrast the differences between early- and late-career teachers in performance and in pay*.
    a. Create two figures for comparison: one showing the distribution of performance for each group (early- and late-career teachers), and the other showing the distribution of salary for each group.
        i. You'll first need to decide the cutoff for years of experience and create a new variable indicating which group each teacher belongs in.

Not satisfied? Improve your plots.
Consider revising your plots:

- Different geoms: consider `geom_jitter`, `geom_violin`, `geom_boxplot`, `geom_density`, etc.
- Different aesthetic mappings: incorporate additional variables by mapping them to color, fill, size, or another aesthetic
- Use a facet to incorporate another variable
- Specify colors with `scale_fill_manual` or `scale_color_manual`
- Specify labels (x- and y-axis, title)
- Customize the legend

[2] The purpose of including a sentence or two is to demonstrate how written text can be woven together with charts and tables in a Markdown file.