

# STAT 331 Lab 5

## Instructions

For this lab you will, again, be submitting an HTML markdown document. We will be working with the San Francisco City Salary Data set posted on [www.kaggle.com](http://www.kaggle.com). The data set is posted on our Poly Learn, but you may also just get it straight from Kaggle.

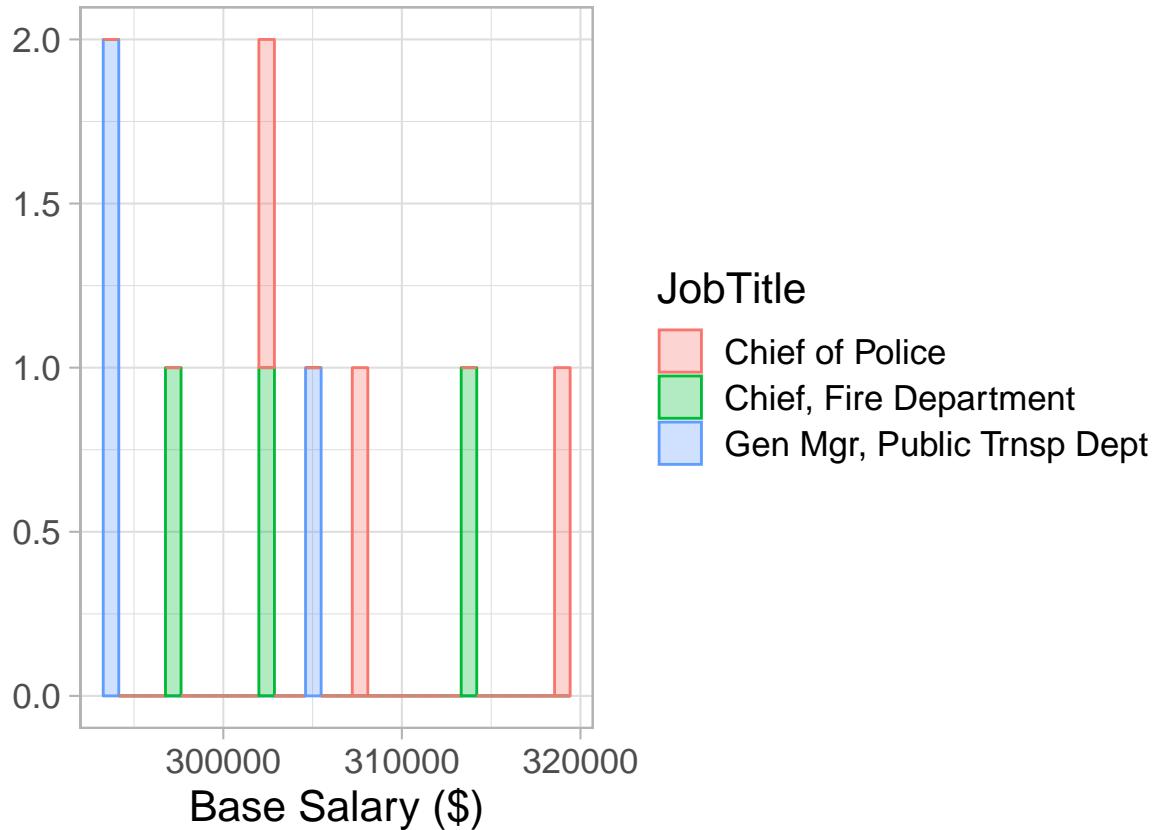
You must use either the *lattice* package or the *ggplot2* package to create the graphs listed below. You may NOT use the Base Plotting System.

Importing the data involves a few special things:

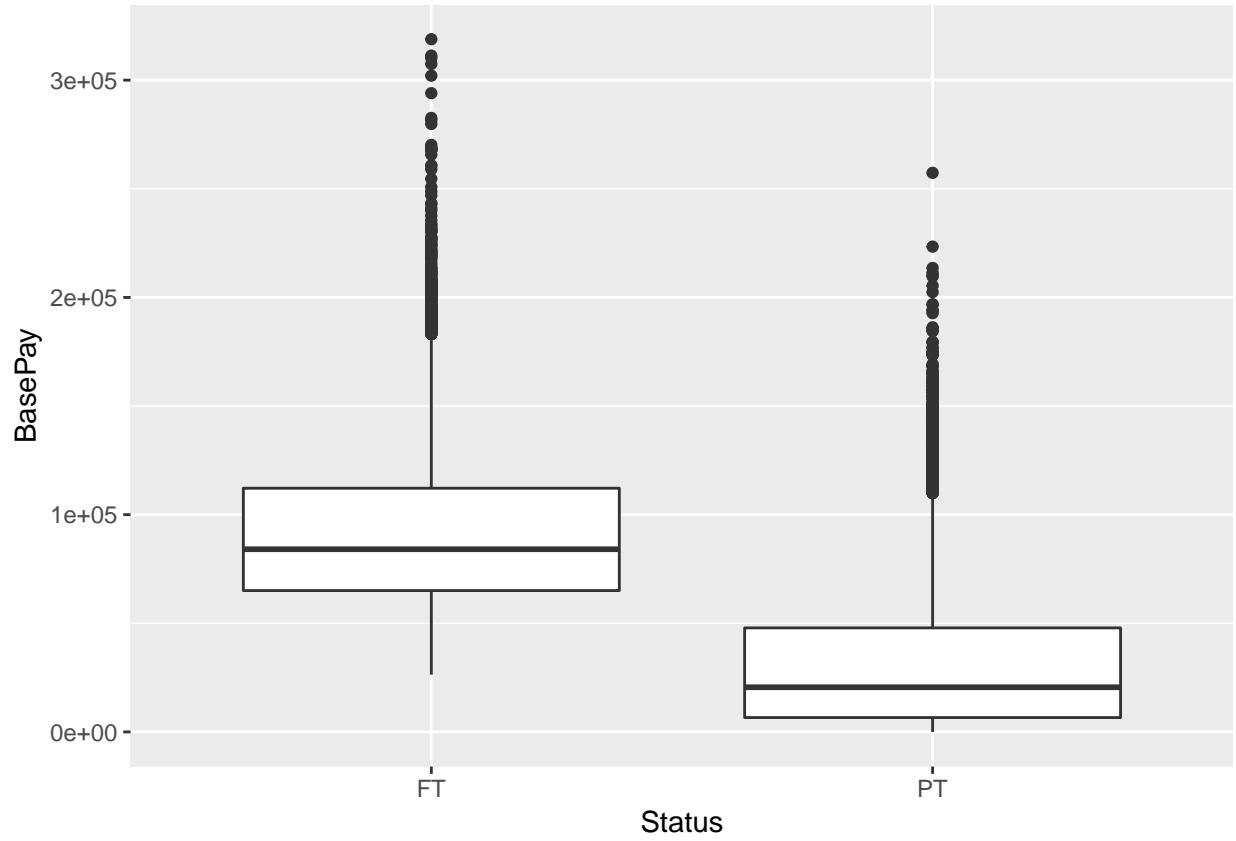
- Make sure `header = TRUE`
- Include `na.strings = "Not Provided"`

## Exercises

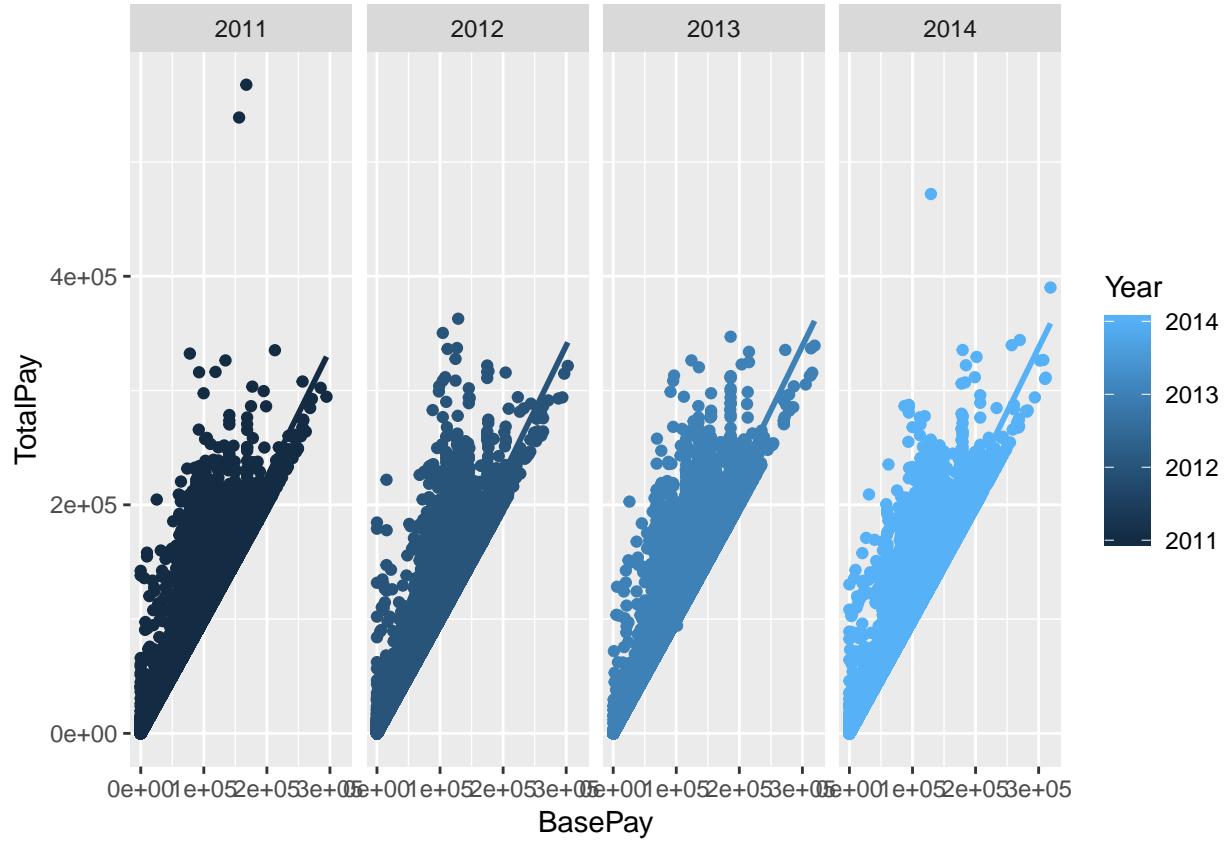
1. Create a plot with overlaid histograms of Base Pay for the 3 jobs with the highest average Base Pay.



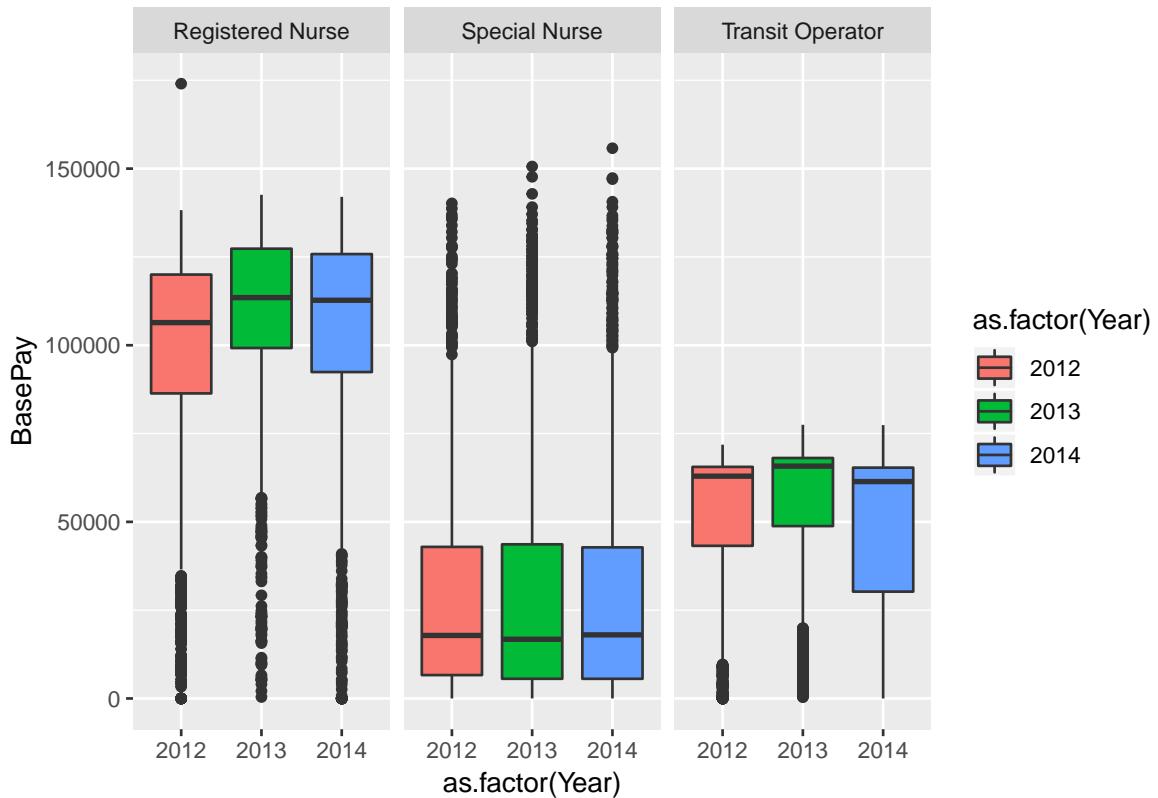
2. Create a plot with side-by-side boxplots of Base Pay by status (FT or PT).



3. Create a multipanel plot (one panel for each value of Year) of Total Pay versus Base Pay with overlaid linear fits.



4. Create a multipanel plot (one panel for each Job) of side-by-side boxplots (one boxplot for each value of Year) of BasePay for the 3 jobs that occur most frequently in the data set.



## Additional Notes

- Be sure to add appropriate titles and axes labels to all of your graphs.
- Make your graphs colorful!
- Consider adding any other details you think would add to the visual appeal of your graphs without taking away from the data.