



Jeffrey Nicolich, Ph.D.

# Introduction

- City of Boston publishes detailed earnings for city employees
  - How is the \$1.7B payroll being spent?
  - Are employees paid fairly (equally)?
  - Is there evidence for wealth segregation?



## Earnings categories

~ 20,000  
employees  
per year

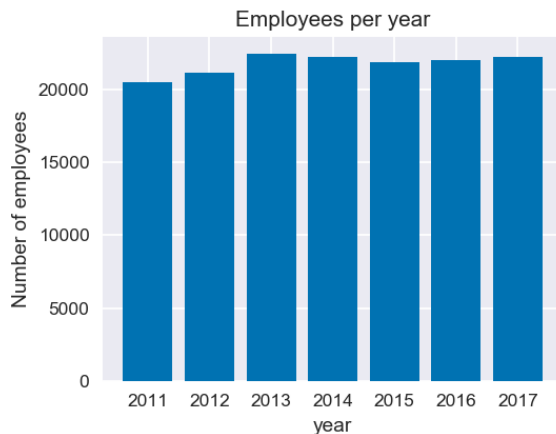
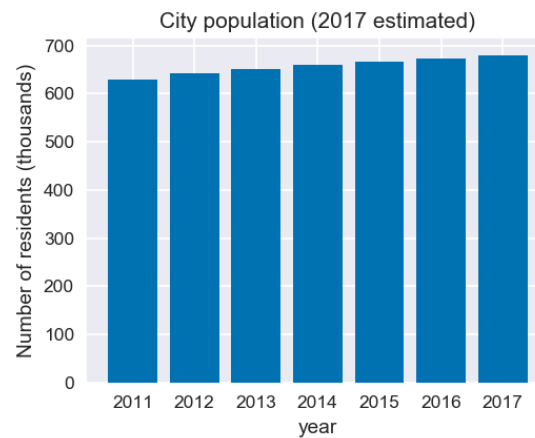
name	department	title	regular	retro	other	overtime	injured	detail	quinn	total	zip	year
Menino, Thomas M.	Mayor's Office	Mayor	175000.00	0.0	0.0	0.0	0.0	0.0	0.0	175000.00	02136	2011
Menino, Thomas M.	Mayor's Office	Mayor	175000.02	0.0	0.0	0.0	0.0	0.0	0.0	175000.02	02136	2012
Menino, Thomas M.	Mayor's Office	Mayor	175000.02	0.0	0.0	0.0	0.0	0.0	0.0	175000.02	02136	2013
Menino, Thomas M.	Mayor's Office	Mayor	10769.23	0.0	0.0	0.0	0.0	0.0	0.0	10769.23	02136	2014
Walsh, Martin J.	Mayor's Office	Mayor	164903.87	0.0	0.0	0.0	0.0	0.0	0.0	164903.87	02125	2014
Walsh, Martin J.	Mayor's Office	Mayor	181730.79	0.0	0.0	0.0	0.0	0.0	0.0	181730.79	02125	2015
Walsh, Martin J.	Mayor's Office	Mayor	175000.02	0.0	0.0	0.0	0.0	0.0	0.0	175000.02	02124	2016
Walsh, Martin J.	Mayor's Office	Mayor	175000.02	0.0	0.0	0.0	0.0	0.0	0.0	175000.02	02124	2017

Years 2011  
through 2017

- Additionally, individual schools are listed for select years
- Data cleaning was needed for a few items:
  - Some departments consolidated and/or changed names
  - Titles are spelled out in various ways
  - Some employees list work place zip code or wrong/missing zip code

# General Trends

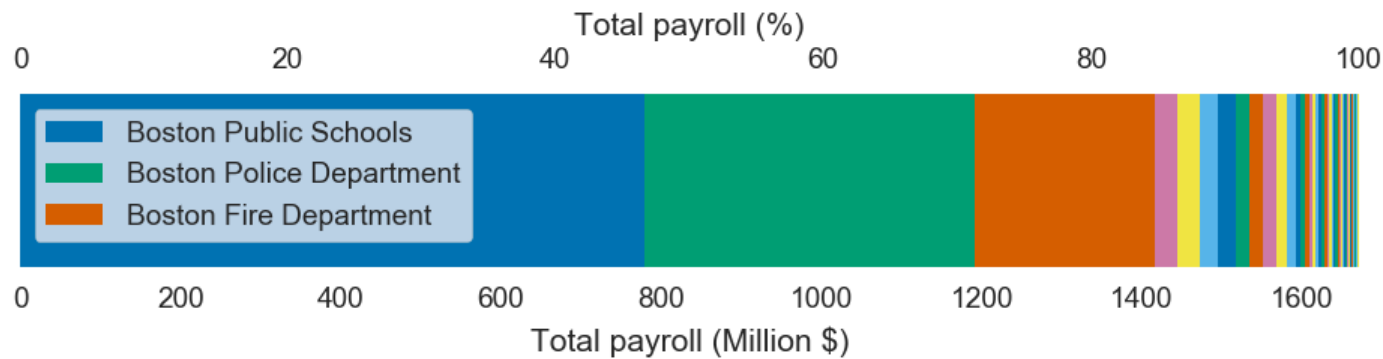
- Biggest increase in total payroll between 2013 and 2014
- Number of employees decreased during same period
- Significant growth in average earnings per employee



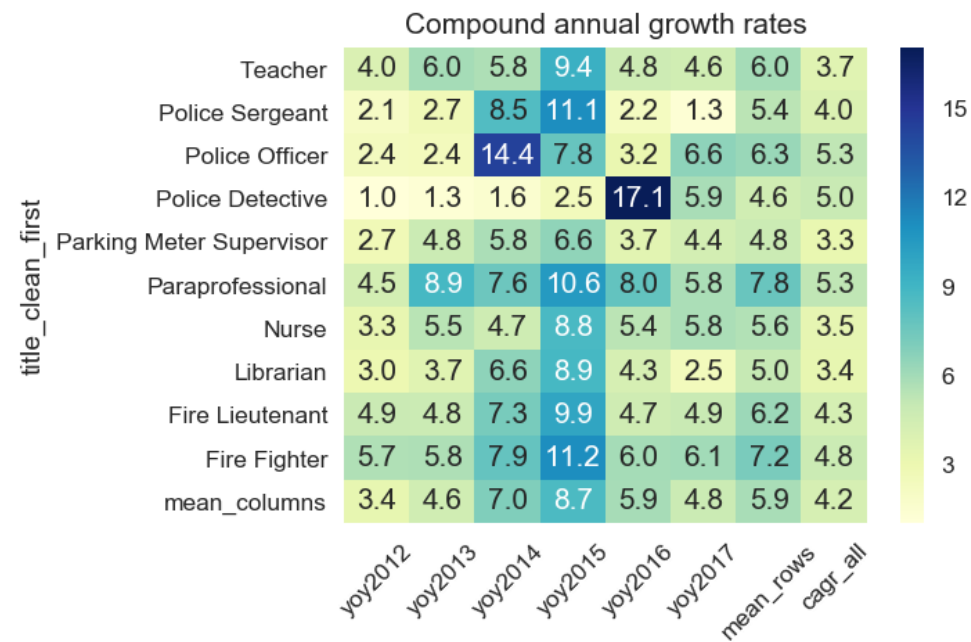
Slower growth than  
SSA average wage  
index in recent years

# Where is the money spent?

- Largest 3 departments account for 84% of payroll



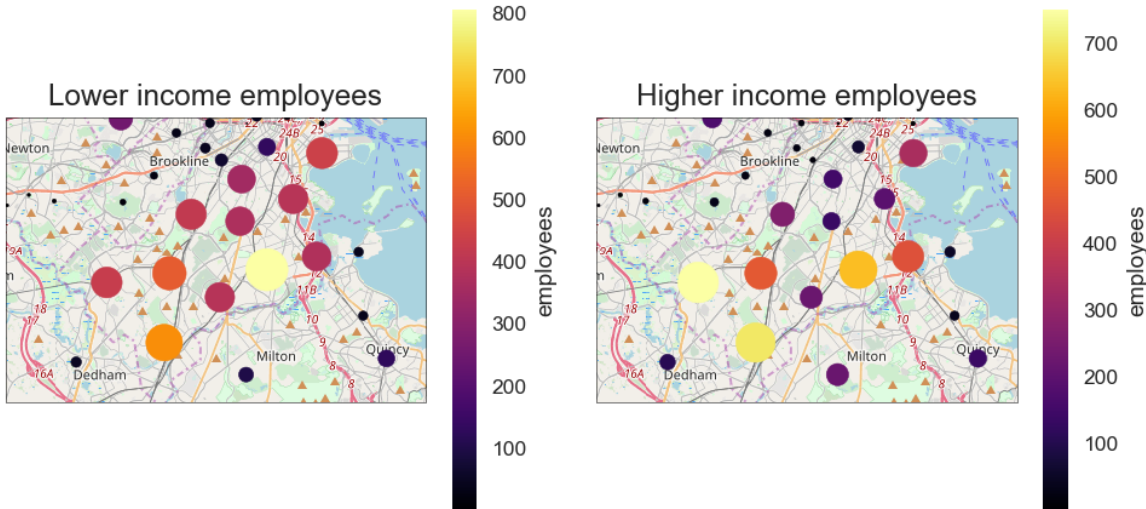
- Pay increases for the 10 most common job titles (excluding partial year employees)
- Increases are uneven over the years
- Growth tends to even out in the long run (“cagr\_all”)
- Largest mean growth in 2015





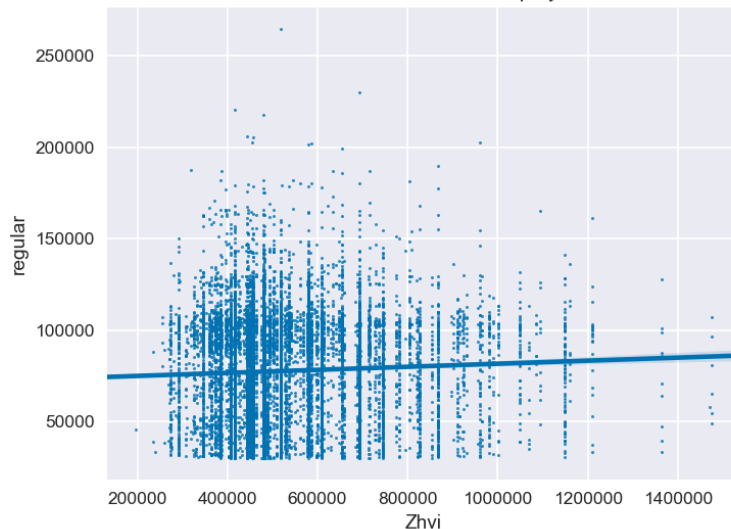
# Wealth Segregation

Wealth segregation (equal populations)

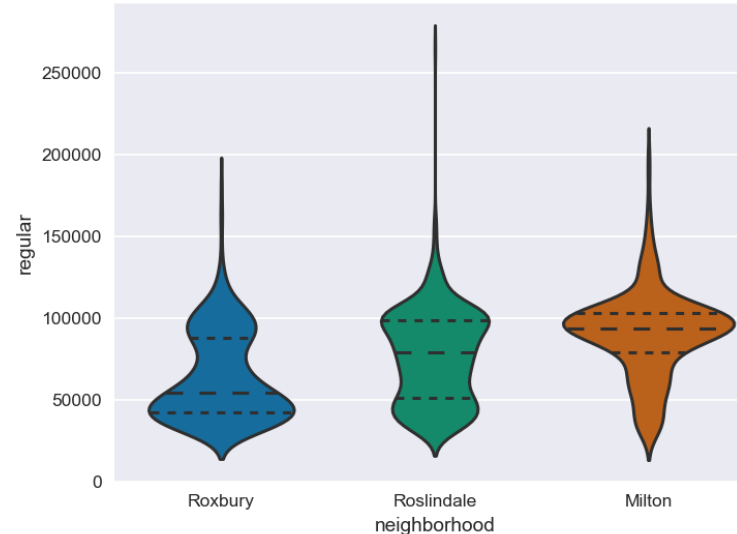


- Wealth segregation can be demonstrated for city employees
- “Zillow Home Value Index” is used as measure of “wealth” per zip code
- Correlation is very weak for all zip codes, but strong for locally well-known extreme cases

Does income influence where employees live?

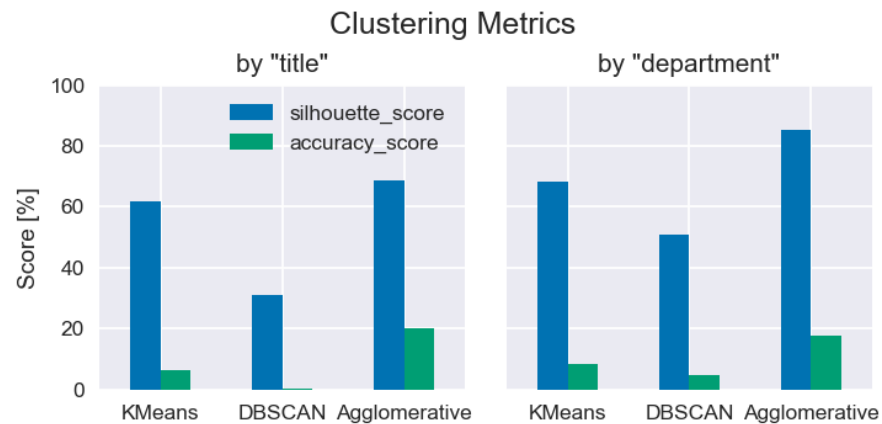


Wealth segregation for select neighborhoods

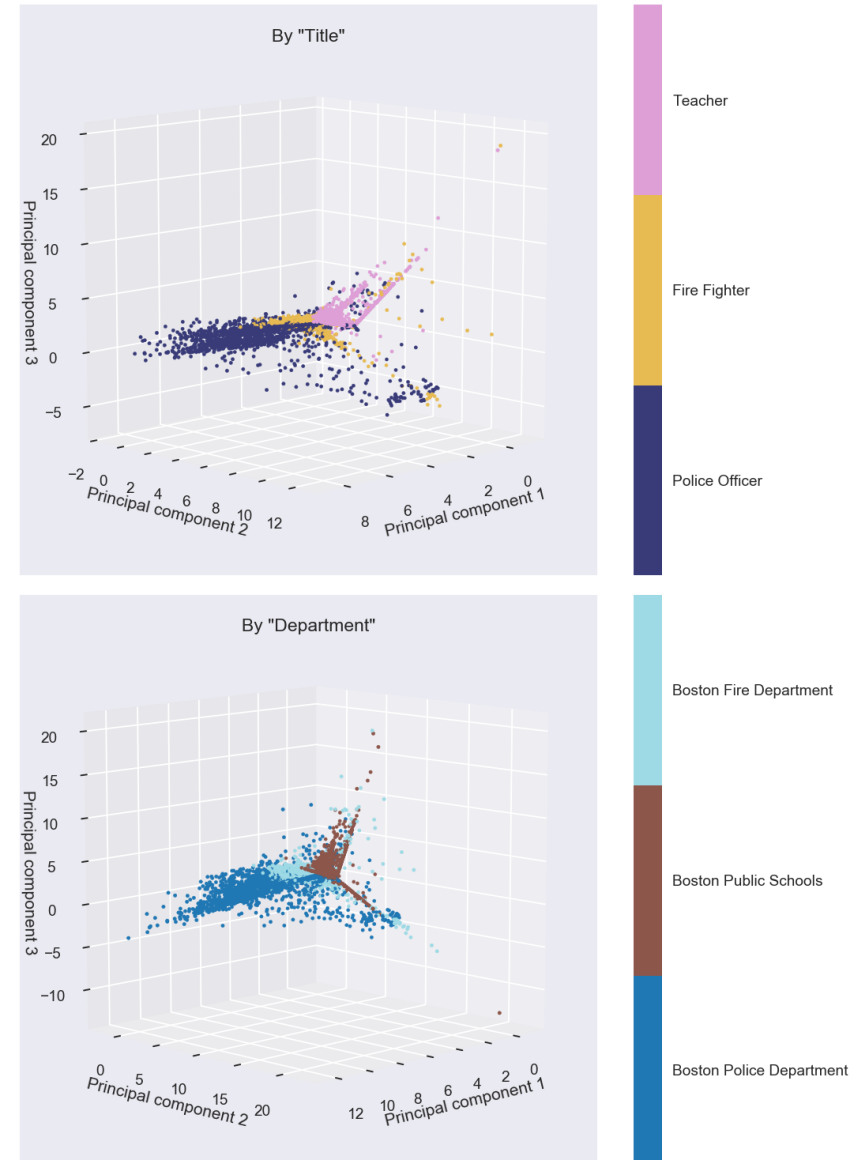


# Pay Discrepancy?

- Can machine learning algorithms predict top 3 job titles or departments?
- Success would indicate significant pay discrepancy
- Unsupervised clustering (cluster analysis) works best with agglomerative clustering algorithm
- Clustering fails to show meaningful structure or prediction accuracy (score)



3D Principal Component Plots



# Pay Discrepancy?

- Supervised machine learning becomes a challenging multi-class classification problem.
- Random Forrest and Support Vector Classifiers work reasonably well if only the most frequent titles or departments are considered.
- Insufficient features (earnings categories) for general classification (all titles/departments)
- “Police Sergeant” is the most misclassified title due to similarity to “Police Officer” and low observation count.
- “Public Facilities Department” is the most misclassified department due to low count in test set and part-time workers.

## Classification by Job Title

	Confusion Matrix				Stats	
	0	1	2	3	count	score
Fire Fighter	340	2	0	16	358	94.97
Police Officer	3	556	0	17	576	96.53
Police Sergeant	1	7	72	1	81	88.89
Teacher	0	4	0	1614	1618	99.75
	0	1	2	3	count	score

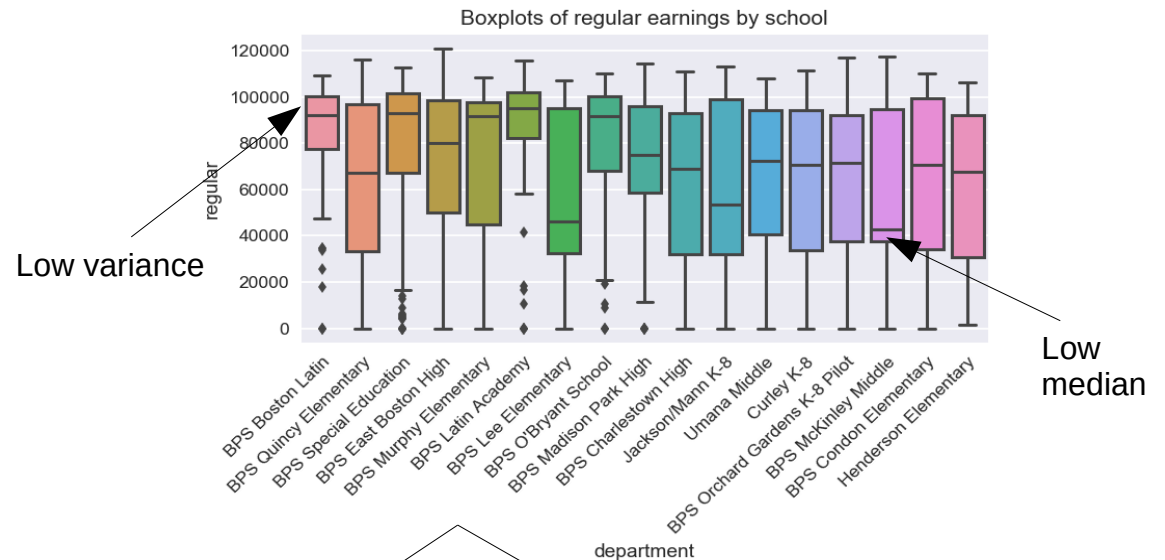
## Classification by Department

	Confusion Matrix					Stats	
	0	1	2	3	4	count	score
Boston Fire Department	283	8	23	0	11	325	87.08
Boston Police Department	22	510	98	0	9	639	79.81
Boston Public Schools	9	10	2608	0	1	2628	99.24
Public Facilities Department	4	10	20	9	1	44	20.45
Public Works Department	2	3	6	0	67	78	85.9
	0	1	2	3	4	count	score

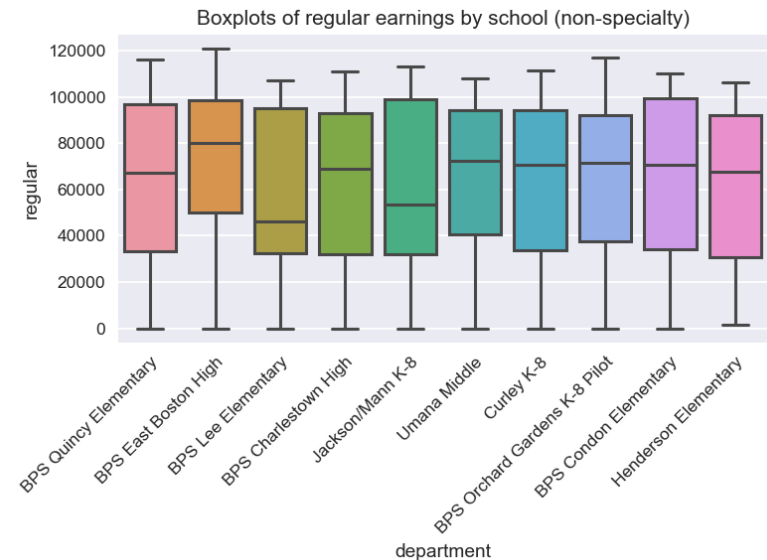
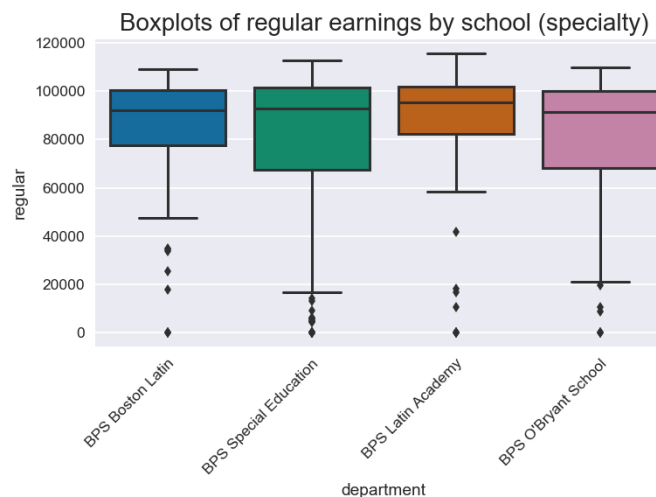
# Are Teachers Paid Equally?

- KNN classification of teacher regular earnings returns an average F1 score of only 14%

- Poor classification results imply no difference in earnings
- Multi-class classification perhaps not the best tool for the job



- Splitting schools into **specialty** and **non-specialty** categories returns very similar distributions within each group





# Are Teachers Paid Equally?

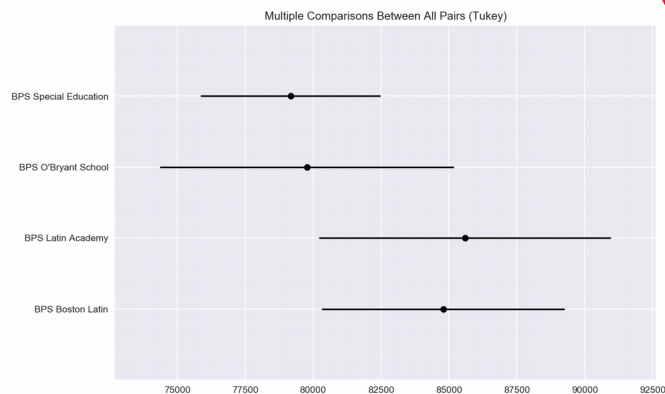
- ANOVA is a useful tool for comparison of means
- Null hypothesis (equal means) is not rejected (reject = False)

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	lower	upper	reject
BPS Boston Latin	BPS Latin Academy	799.5256	-9073.7494	10672.8005	False
BPS Boston Latin	BPS O'Bryant School	-5020.1532	-14924.7595	4884.453	False
BPS Boston Latin	BPS Special Education	-5618.4619	-13321.7146	2084.7908	False
BPS Latin Academy	BPS O'Bryant School	-5819.6788	-16526.6053	4887.2477	False
BPS Latin Academy	BPS Special Education	-6417.9875	-15128.7367	2292.7617	False
BPS O'Bryant School	BPS Special Education	-598.3087	-9344.5546	8147.9373	False

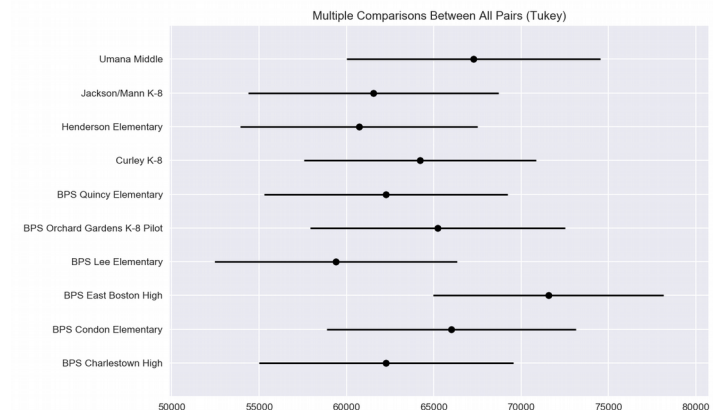
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	lower	upper	reject
BPS Charlestown High	BPS Condon Elementary	3747.849	-10647.9868	18143.6849	False
BPS Charlestown High	BPS East Boston High	9295.4976	-4578.4972	23169.4924	False
BPS Charlestown High	BPS Lee Elementary	-2879.9703	-17085.454	11325.5134	False
BPS Charlestown High	BPS Orchard Gardens K-8 Pilot	2953.1201	-11616.408	17522.6481	False
BPS Charlestown High	BPS Quincy Elementary	-19.7135	-14255.6638	14211.2368	False
BPS Charlestown High	Curley K-8	1951.0158	-11974.2099	15876.2416	False
BPS Charlestown High	Henderson Elementary	-1542.2513	-15602.3146	12517.812	False
BPS Charlestown High	Jackson/Mann K-8	-710.4463	-15139.8597	13718.967	False
BPS Charlestown High	Umana Middle	5001.2926	-9532.3055	19534.8908	False



Specialty

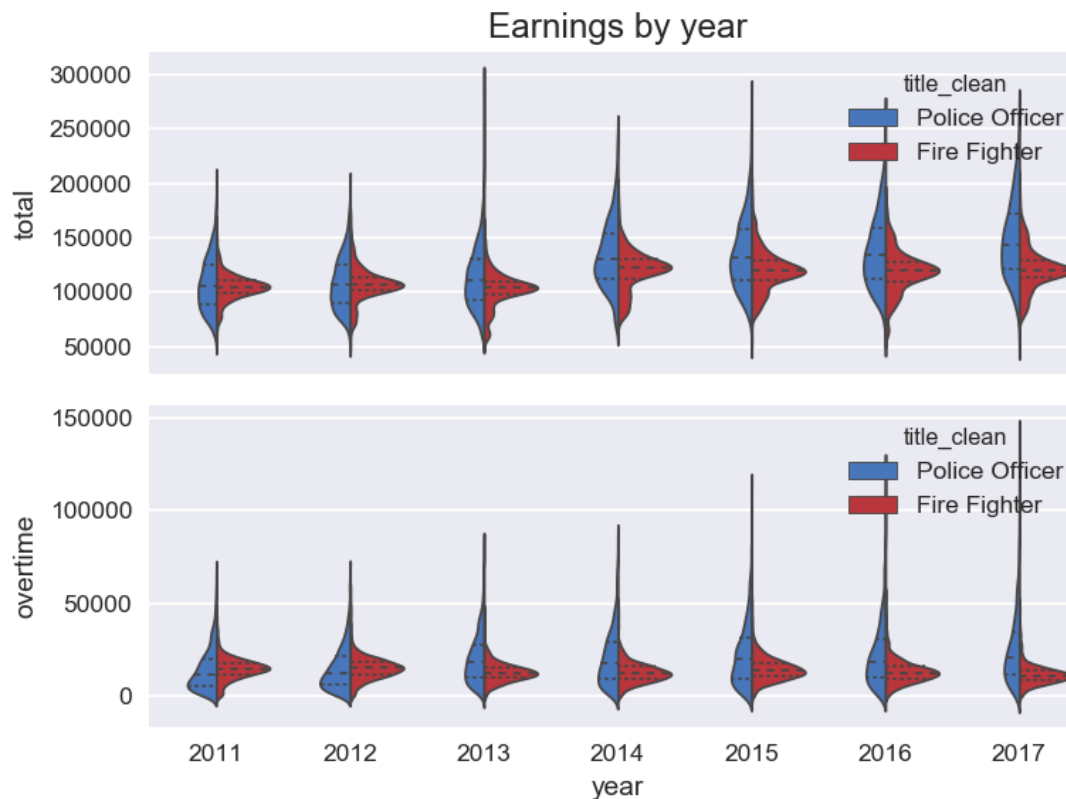
Non-specialty



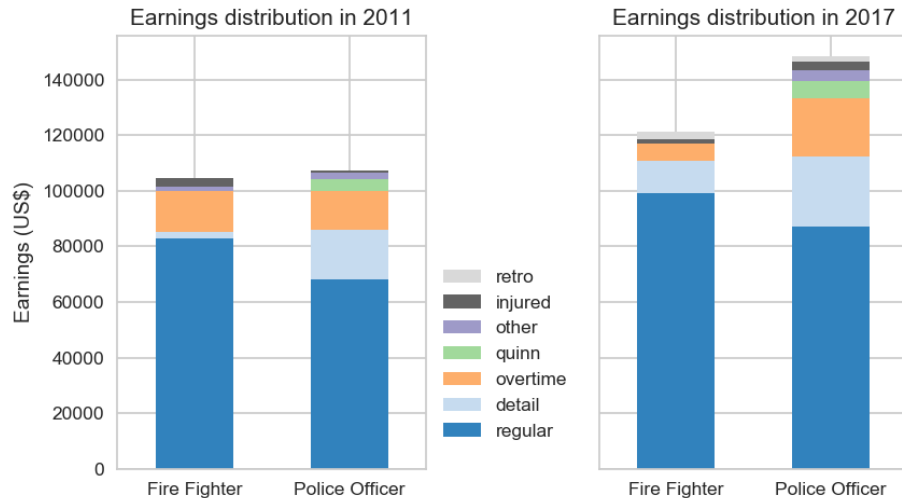
- Magnet schools and other specialty schools have lower variance
- Low variance could be result of lower turnover
  - More desirable workplace
  - Specialized education or job description
  - Higher pay, but not necessarily
- Key information to measure turnover is missing from the dataset

# Police Officers and Firefighters

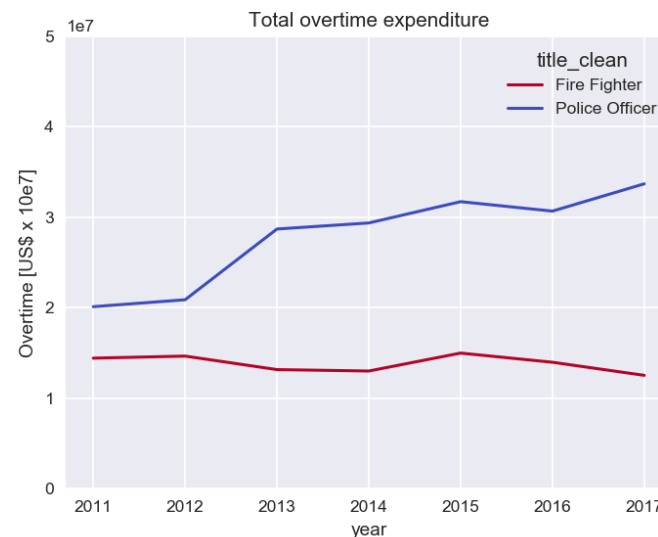
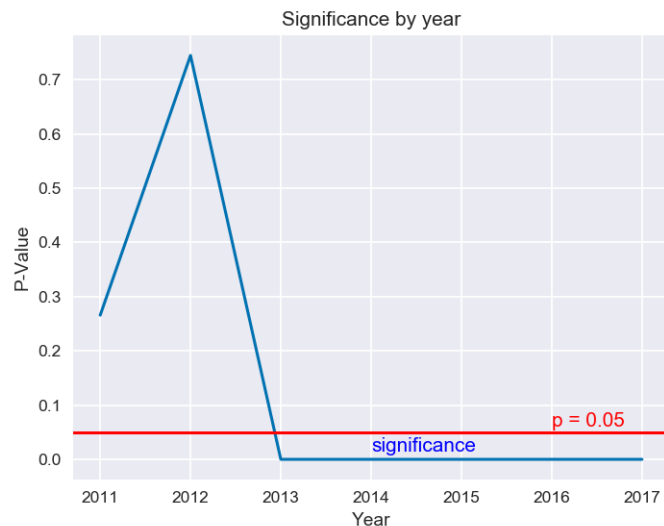
- In 2011, firefighters have slightly higher median total earnings
- By 2017 police officers earn more
- Biggest change is overtime pay



# Police Officers and Firefighters



- In 2011 and 2012, there was no measurable difference in mean overtime pay between firefighters and police officers
- Total overtime pay for police officers is \$13 million above 2011 level
- Slight decrease in overtime for firefighters



# Conclusions

- No convincing evidence of pay inequality for same or similar titles
- Total payroll increase is driven by increased overtime, in particular from police department
- Wealth segregation is evident, but mitigated by a variety of factors beyond the reach of the earnings database.
- Better conclusions could be drawn if additional information were available:
  - Employment start date
  - End date
  - Promotion date
  - Annual salary or expected earnings



# References

- Main dataset:
  - <https://data.boston.gov/dataset/employee-earnings-report>
- Boston public schools department
  - <https://www.bostonpublicschools.org/>
- Zillow Home Value Index:
  - <https://www.zillow.com/research/data/>
- Social Security Average Wage Index:
  - <https://www.ssa.gov/oact/cola/awidevelop.html>
- Population Census:
  - <https://www.census.gov/quickfacts/fact/table/bostoncitymassachusetts/PST045216>
  - 2017 data is extrapolated from 2011-2016 data
- Commentary on police department overtime hours:
  - <https://www.bostonglobe.com/metro/2018/02/16/bpd-captain-was-city-top-earner>