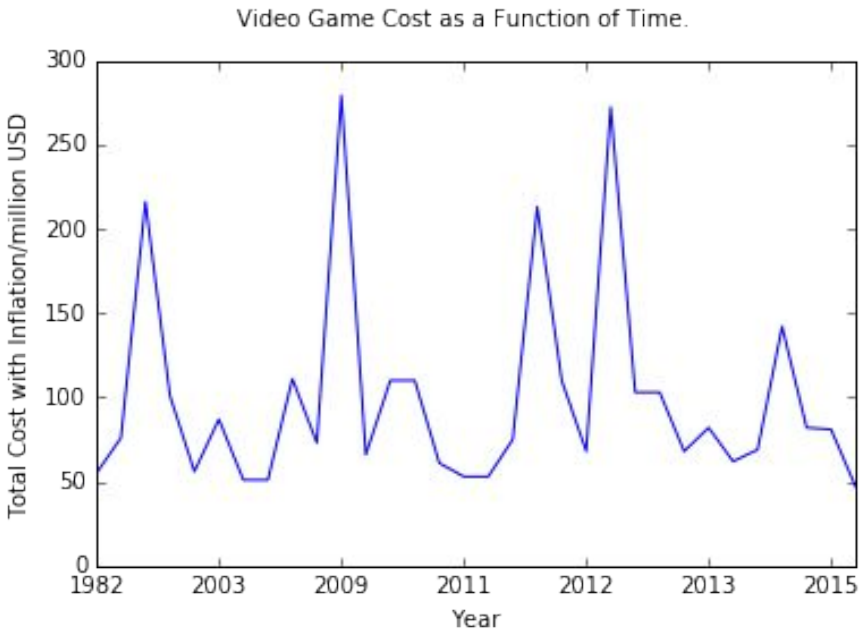
# Project 2

Is salary related to experience for my kids's teachers?

Chris Harwell

# Implementation



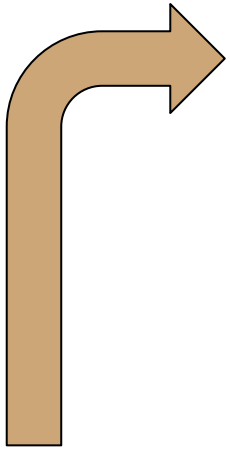

- Recall I started with looking at expense of the most costly to produce video games
- Pulled from wikipedia table
- Found some others to cross-reference
- But just not enough data points with consistent, reliable info
- Burned a bunch of limited time, but ultimately moved on to a topic more interesting and closer to home ….

# DATA UNIVERSE

Powered by the Asbury Park Press, part of the USA TODAY NETWORK

**Start searching millions of public records**

Select a school, enter a name or a job title to search for public school educators certified by the state. The list is released once a year by the state Department of Education and provides information about jobs, salaries, types of degrees and years as a teacher as of Oct. 15, 2016….



NJ Teacher Salaries - U

php.app.com/agent/educationstaff/search/page:34/sort:salary/direction:asc?last_name=&first_name=&county=MORRIS&distr

Apps | DataSciProj2 | hackercode-refs | Unemployment | DataSciIdeas | DataSciRefsReading | Imported From Fire | Kindle Clo

DATA UNIVERSE

Payroll | Homes & Taxes | Schools | Crime | Retirees & Other | Statistics | Subscribe

30 31 32 33 34 35 36 37 38

Page 34 of 42, showing 10 records out of 417 total, starting on record 331, ending on 340

| First Name | Last Name | County | District | School | Job | Salary |
|---|---|---|---|---|---|---|
| Ronald | Deloatch | Morris | Sch Dist Of The Chathams | Chatham High School | Social Studies History | $87,475 |
| Gail | Hatch | Morris | Sch Dist Of The Chathams | District Office | Physical Therapist | $88,150 |
| James | Miller | Morris | Sch Dist Of The Chathams | Chatham High School | Science Earth | $88,150 |
| Cynthia | Gagliardi | Morris | Sch Dist Of The Chathams | Chatham High School | English Non-Elementary | $88,150 |
| Oona | Abrams | Morris | Sch Dist Of The Chathams | Chatham High School | English Non-Elementary | $88,410 |
| Cindy | Weiner | Morris | Sch Dist Of The Chathams | Lafayette Avenus School | School Counselor | $88,410 |
| Bina | Patel | Morris | Sch Dist Of The Chathams | District Office | Occupation Therapist | $89,045 |
| Dorothy | Mccorry | Morris | Sch Dist Of The Chathams | Milton Avenue School | Elementary Kindergraten-8 Grade | $89,285 |
| Carolanne | Carty | Morris | Sch Dist Of The Chathams | Southern Boulevard School | Resource Program Pull-Out Support | $90,495 |
| Dean | Kravitz | Morris | Sch Dist Of The Chathams | Lafayette Avenus School | Music Comprehensive | $90,517 |

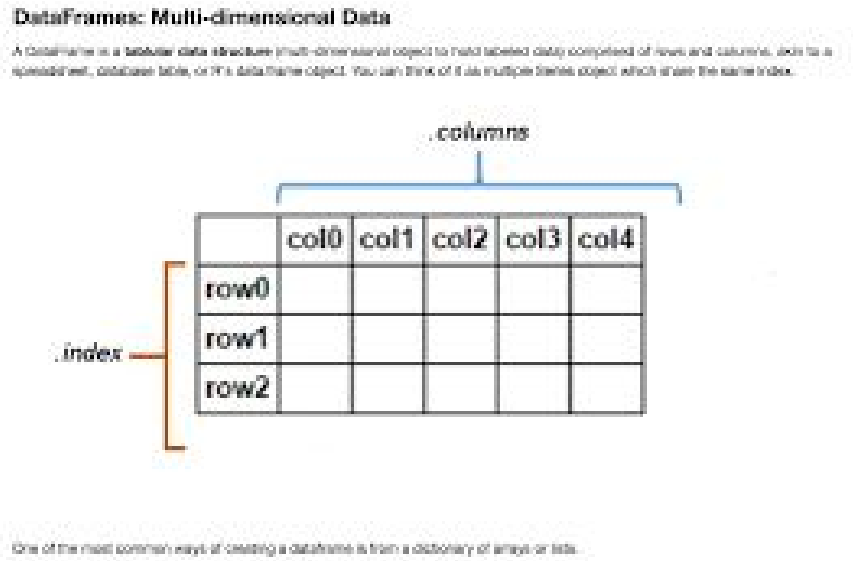
County: MORRIS

District: All

School: All

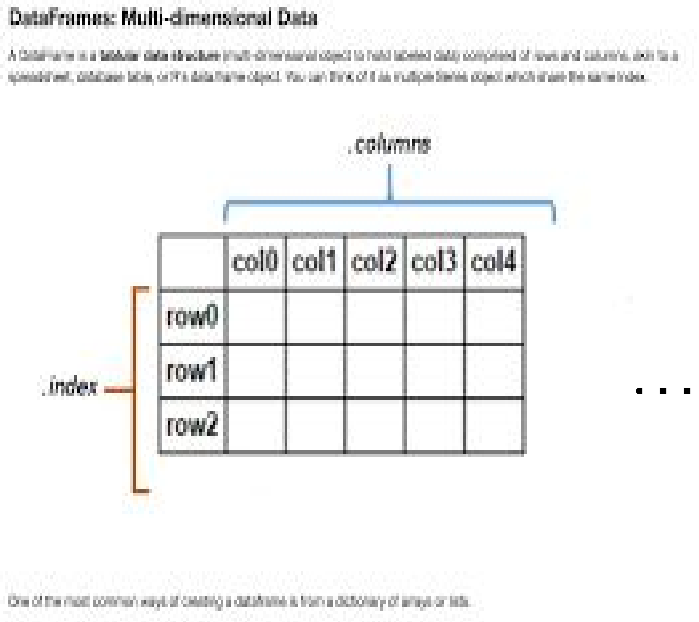
Reset Search





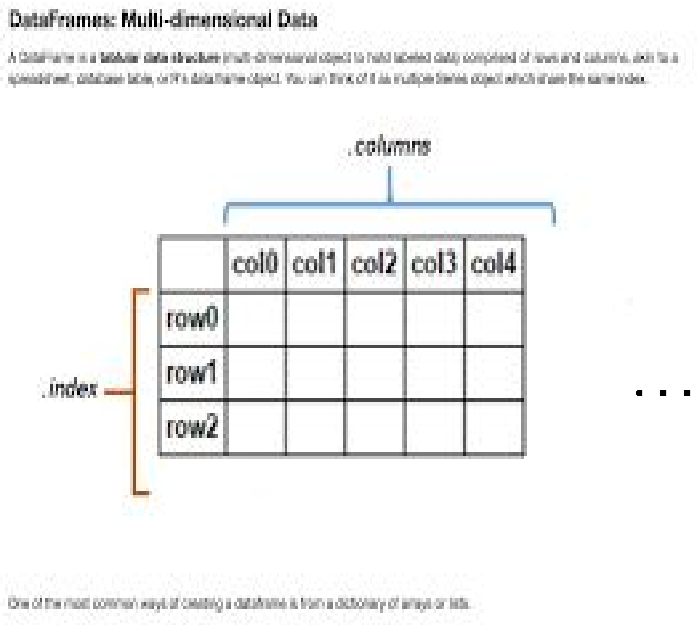

Python 3.5.3
Python3-requests-2.10.0-4.fc25.noarch
Python3-beautifulsoup4-4.6.0-1.fc25.noarch
python3-pandas-0.19.0-1.fc25.x86_64

Pages in biglist: 13,873
Narrowed down to employee id range: 92545-91953
Morris Cty 592 -> filter -> Chatham: 396




row * column: (393, 15)

1 column: first
2 column: last
3 column: ***salary ~ f(x_1 …x_n)?***
4 column: county
5 column: district
6 column: *experience_district*
7 column: school
8 column: experience_nj
9 column: primary_job
10 column: experience_total
11 column: fte
12 column: subcategory
13 column: certificate
14 column: highly_qualified
15 column: teaching_route




row * column: (393, 15)

1 column: first
2 column: last
3 column: ***salary ~ f(x_1 …x_n)?***
4 column: county - constant
5 column: district - constant
6 column: ***experience_district***
7 column: school - constant
8 column: ***experience_nj***
9 column: primary_job
10 column: ~~***experience_total***~~
11 column: ***fte***
12 column: **subcategory**
13 column: certificate
14 column: highly_qualified
15 column: teaching_route

# Salary


average: 71,931

middle: 65,035

min,max range: 21,236 through 164,303

iqr range for dispersion: 22,884

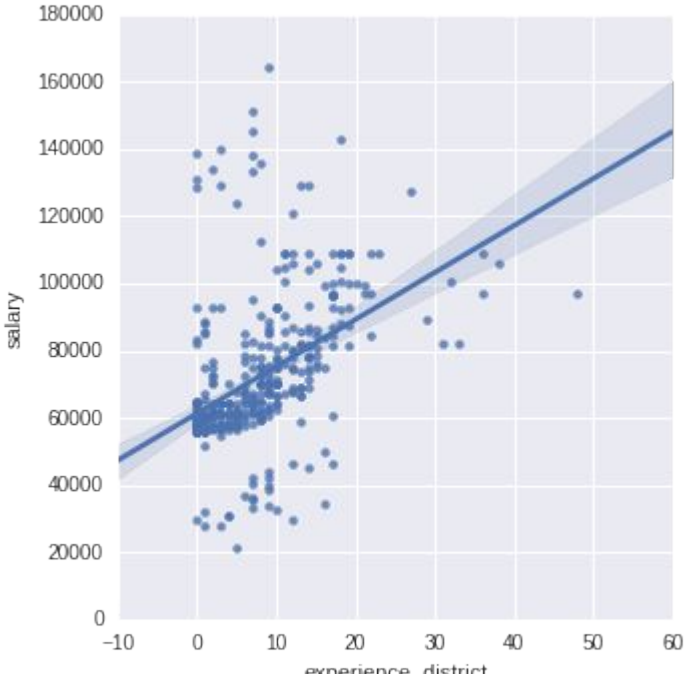
# Is district experience enough?

Modeling salary as a function of district experience alone only accounts for about 20% of the variability.

So, include more … like experience within NJ and full-time equivalency.

Salary as a function

of only district experience



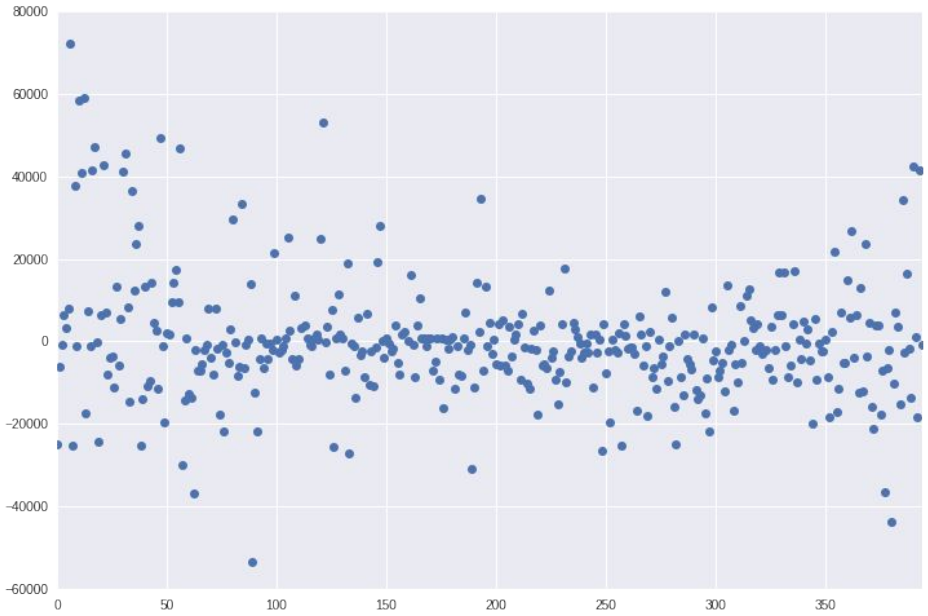
# salary function of ... experience_district + experience_nj + fte....?

R^2 = 0.54 (less w/o fte)

Salary ~=
- 1209 * experience_district
+ 2781 * experience_nj
+ 37530 * fte
+ 17,960

Independent variable: y-axis salary

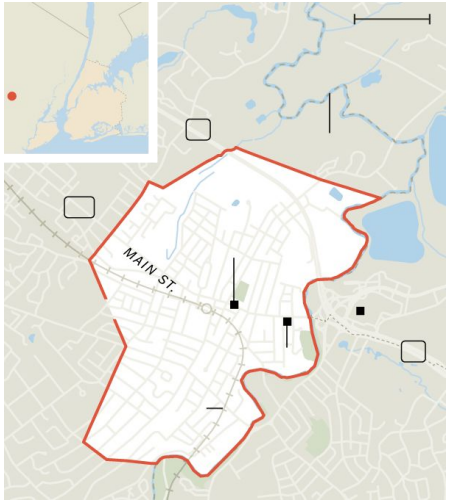
Slight bow to residuals - from 1.2 skew?



- Including subcategory -- such as general vs. special education vs. administration -- accounts for additional variability bringing R-squared to 0.758,

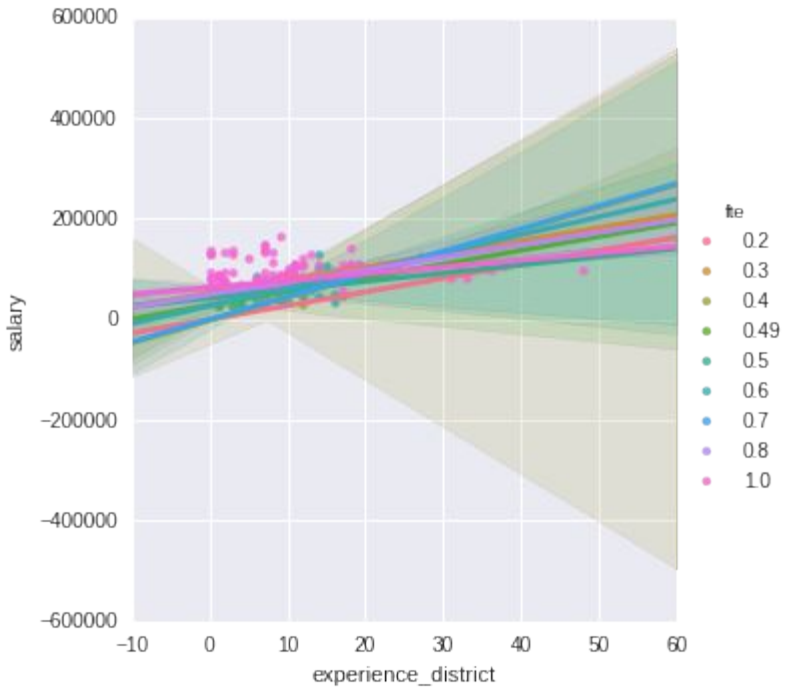
- Subcategory:
  - General ed 318
  - Special ed 56
  - Admin or supervisor 18
  - Hearing 1

- ⅔:⅓ test/train split is low at R-squared 0.414

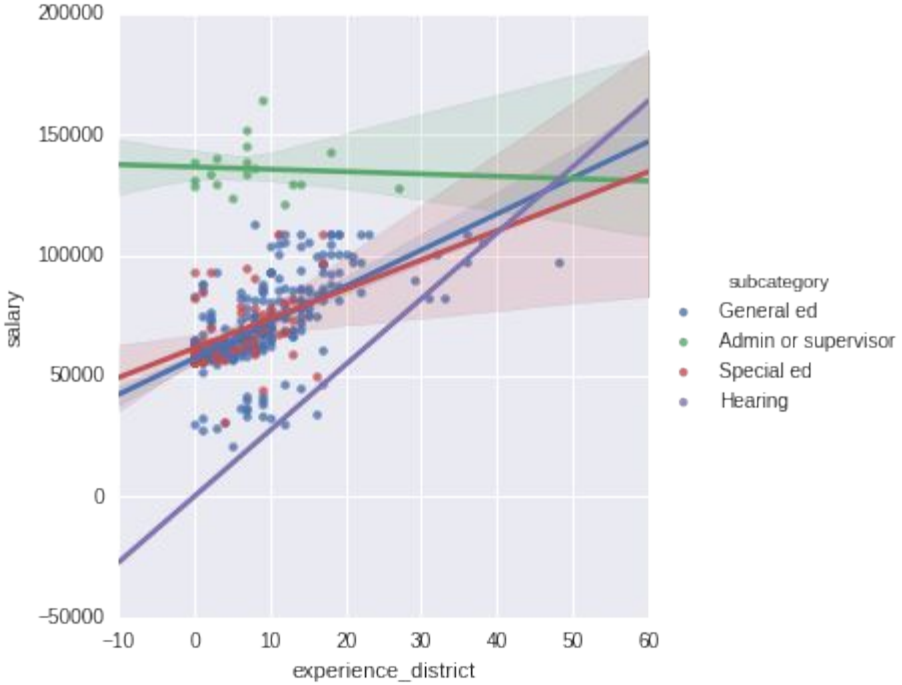


- Might be fun to also include:
  - other districts, but adjust with COLA
  - group by union/non
    - Subject
    - grade level
  - include categorical info such as certificate, qualification, highest degree, teaching route
  - Demographics
- More model evaluation.
- Try median with a quantile regression (vs. mean here with least squares).

Chris R. Harwell
chris.r.harwell @gmail.com
https://www.linkedin.com/in/chris- r- harwell
Chatham, NJ mobile: (862) 246-6142




Salary as a function of district experience, with hue for full time equivalence..




Salary as a function of district experience, with hue for subcategory - administration vs. general teaching vs. special.

harwell_chris_project_ ×

Chris

localhost:8888/notebooks/harwell_chris_project_2_data_universe_edu_sal_regress

Apps | DataSciProj2 | hackercode-refs | Unemployment | DataSciIdeas | » | Other bookmarks

jupyter harwell_chris_project_2_data_universe_edu_sal_regressions (unsaved changes)

File Edit View Insert Cell Kernel Help | Python 3

Code | Cell Toolbar: None

In [29]:
```
y, X = patsy.dmatrices('salary ~ experience_district + experience_nj + fte', data=df
model = sm.OLS(y, X)
fit = model.fit()
fit.summary()
```

Out[29]:

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | salary | **R-squared:** | 0.539 |
| **Model:** | OLS | **Adj. R-squared:** | 0.535 |
| **Method:** | Least Squares | **F-statistic:** | 151.4 |
| **Date:** | Fri, 14 Jul 2017 | **Prob (F-statistic):** | 5.09e-65 |
| **Time:** | 12:53:31 | **Log-Likelihood:** | -4331.0 |
| **No. Observations:** | 393 | **AIC:** | 8670. |
| **Df Residuals:** | 389 | **BIC:** | 8686. |
| **Df Model:** | 3 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[95.0% Conf. Int.]** |
|---|---|---|---|---|---|
| **Intercept** | 1.796e+04 | 4761.009 | 3.771 | 0.000 | 8595.270 2.73e+04 |
| **experience_district** | -1209.2907 | 209.348 | -5.776 | 0.000 | -1620.886 -797.695 |
| **experience_nj** | 2781.8870 | 185.936 | 14.962 | 0.000 | 2416.322 3147.453 |
| **fte** | 3.753e+04 | 4791.543 | 7.833 | 0.000 | 2.81e+04 4.7e+04 |

# Ordinary least squares regression results