Philadelphia: A Big American City on The Verge of Booming

Mid-term Report

Michael Uftring, Indiana University

Abstract. The City of Philadelphia has a checkered past, both economically and socially. Twentieth century history depicts Philadelphia as a city very much on the decline in almost all measured socioeconomic categories. Entering the Great Recession of 2007-2009, Philadelphia was in one of its worst declines with high crime rates, widespread poverty, a collapsing housing market, and an almost nonexistent job market. Since the end of the Great Recession, Philadelphia has started a major turnaround. The city has developed and promoted great leadership (through both elected officials, and community volunteers), initiated social, cultural and economic development programs, become visible on the world stage, and continues to find ways to remake itself into a notable American city. Philadelphia is poised to move into the top ranks in all socioeconomic growth and development categories and be a major contributor to the future of The United States of America .

Introduction

Philadelphia: commonly known by nicknames like The Birthplace of America, and The City of Brotherly Love. American history books cannot be written without significant mention of Philadelphia's place in the creation of The United States of America. However, Philadelphia, like most American cities, has experienced an up-and-down history; recently seeming to know far fewer ups than downs.

The Great Recession of 2007-2009 took a major toll across all of America; cities and rural areas were all affected. The recession did not spare any industry, race, gender, or socioeconomic status. Philadelphia is rebounding from the recession, but the city is not content to just get back to business as usual per the pre-recession status and statistics. This research project aims to show how The City of Philadelphia is making and executing plans to become The Next Great American City.

The report will examine areas including:
- Economic growth, in terms of industry and jobs
- Social growth and improvement; education, equal opportunity, diversity
- Cultural impact; in the arts, and with a national and global presence

This report will show how Philadelphia is doing compared to:
- National indicators

- Regional indicators
- Other cities with similar demographics (population, geography, ethnicity)

A Compelling Interest

This project is important because it aims to show that an old industrial city, cast off by almost everyone and in deep decline, can make a comeback and be relevant and become a role model. But, can these old cities really rise above being relics and artifacts of Colonial American history? There are many examples: Philadelphia, Detroit, Cleveland, St. Louis. All desire to be more than just National Monuments of their respective noteworthy eras. The changes are difficult, and the competition from other cities for industry and jobs is fierce.

I have a deep personal interest in this topic as well, and that was a significant part of the motivation in pursuing this as a research project. I moved to Philadelphia in the midst of The Great Recession, have decided to raise a family here, and desire to live here "forever." As I near completion of my Master's degree in Data Science, I begin to see the change in career trajectory. My next career phase interests align well with Philadelphia's strengths (healthcare, biotech, research, education) and I am seeking to know whether that "live here forever" decision is viable.

The Big City Landscape

Being in the The Big City Landscape is a challenging place for any city to thrive, big or small. There is endless competition for cities to be competitive and be able to attract businesses which will bring jobs and prosperity to their region. An industry, or even a single company, can come to define a city and region. This was true in the industrial era, and holds true today.

The big companies aren't just good for the city, they become good for other companies as well. Their presence spurs and incubates competition from small startups to "compete" with the bigger companies, or "fill the gaps" not satisfied by the bigger companies. All of this creates an environment for a strong economic ecosystem that provides long-term growth and prosperity to a city and its surrounding region.

Even once a company decides to call some city home, there's nothing stopping other cities from trying to lure the company away. And in some cases, a company becomes discontent with their home city and opts to relocate. Because of this facet of the competitive and somewhat fickle environment, every city must become an attractive destination and continue to change and improve itself to remain attractive.

Philadelphia

Philadelphia's strengths have long been healthcare and higher education [shrm]. The major universities located in the city -- The University of Pennsylvania (Penn), Temple University, Drexel University, and Jefferson University -- all have very strong, world renowned medical programs, and each have an

affiliated teaching hospital. At least two have spun off self-contained, thriving healthcare systems: Penn Medicine, and Jefferson Healthcare.

A physical indicator of progress and economic growth must be large-scale construction, both commercial and residential. Cranes, scaffolding, and iron workers have been a steady presence in the Philadelphia skyline for several years. One can look into Center City, the traditional locus of business, and see many buildings redefining the skyline. And in the near part of West Philadelphia, in an area dubbed *University City* because of the large defining presence of Penn and Drexel, tall buildings are shooting up in an area once just made up of low-rise buildings; this is creating what locals are calling The Second Skyline. [curbed-1, curbed-2]

Construction (in progress, or recently completed)
- Comcast Technology Center
- Cira Center South (FMC Tower)
- Aramark
- Penn Medicine
- Children's Hospital of Philadelphia
- Market East (Jefferson Station area)
- The Navy Yard (in South Philadelphia)

Construction (in planning, or starting)
- 30th Street Station District
- Schuylkill Yards
- uCity Square (University City, West Philadelphia)
- Northeast Philadelphia (primarily residential, and transportation related)

A Project in Two Parts

After some initial background material reading, and data exploration and analysis, it became clear that this topic area has the potential to be a very large project. The options became: reduce the scope to make it a smaller project, or divide and conquer. As it happens, both approaches were needed. I didn't want to reduce the scope too much, because I felt it would render the analysis and report rather boring and perhaps useless. Also, I needed to structure the two parts into something which (1) showed progress, (2) could be logically divided, and (3) when combined would produce a meaningful and useful report.

Part I

The first part of this project focused on: data acquisition, data exploration, and the first level analysis on population and employment. In addition, I needed to determine what data and analysis would follow the first part. The "what's next?" question was primarily determined as a result of the first part analysis. This also became the reasoning for dropping certain avenues of analysis.

In the second part I will explore and analyze the "what's next?" direction. This will entail adding industry and jobs as dimensions of data to the analysis. The should show if there is a shift both in and out of Philadelphia of certain jobs and industries. Comparisons will be made with other cities with similar characteristics; the objective being to determine whether there are any similarities and whether they may indicate economic growth.

The ultimate project deliverable is to build a small dashboard tool which will show a few "interesting" visualizations derived from the experience of this project. The tool will permit the user to select a city for the basis of comparison, and use that create the visualizations supporting the two comparison groups: regional, and by population size. For simplicity, the city selection will be restricted to the top 100 from The List of United States cities by population [wikipedia-1].

Finally, I will conclude by providing my own answer to the question: *is Philadelphia going to have a "big boom" in its future?*

Fallen out of Scope

Some of the socioeconomic data is hard to interpret correctly without significant research into the underlying causes and trends. As such, these dimensions of data analysis are being removed from the project scope.

Background and Related Work

Philadelphia: An Incomplete Revival (The CCD Report) [ccd] provides a historical perspective into how Philadelphia got to where it is. It provides a generally bleak picture of the current growth rates and recovery, but also provides a generally positive outlook for the future. Metro Economic Outlook: Philadelphia [shrm] also provides some valuable insight and outlook, but with some inadequate visualizations which are aimed to be improved upon in this report.

News stories are published and broadcast in local, regional, and national media -- Philadelphia Magazine [phillymag], The Philadelphia Inquirer [inquirer], The Philadelphia Citizen [citizen], NewsWorks [newsworks], and NPR [npr-twoway, state-impact]. These provide the city's current "heartbeat" and some perspective into the current perception of progress and growth (or lack thereof). These reports also tend to have a layer of opinion on government policy changes, or business actions that would have a great impact (again, positively or negatively).

The U.S. Census Bureau and "open data" agencies [usa-data, usa-stats], The Bureau of Labor Statistics [bls], and Brookings Institute [brookings] provide the foundational data which will be analyzed and will be the source of key visualizations. The City of Philadelphia [opendata, philadata] and The Philadelphia

Business Journal [philabusj] provide local and regional data which may prove helpful in the interpretation and analysis of the data provided by federal agencies.

Working Hypothesis

The key indicators in some reports show that Philadelphia has been growing at a pace which lags comparative cities and regions [ccd]. This reporting of the past is not disputed. However, these facts and figures cannot take into account intangible indicators of growth like large construction projects which are well into planning, are underway or are near completion. These more recent developments and the prevailing trend will ideally show that Philadelphia is approaching or consistent with traditionally recognized "great cities." The potential in the to-be-realized growth rate will hopefully continue an upward trend, increase its velocity, and project Philadelphia into the upper tier of great economic cities. With this Philadelphia will be a more significant factor in the future U.S. economy and society. Philadelphia could become a destination for diverse industries (not just those traditionally associated with the city), and the city could become a model for rebuilding cities with a tumultuous industrial past.

The key measurable indicators analyzed in this report are:
- Jobs: Unemployment Rates, Labor Force
- Industry and Occupation
- Population

Nationally, population is up and unemployment is down. Is the population level in a city a good indicator of employment and predictor of economic potential or growth? Do the national trends map (somewhat) directly onto or from the city trends?

Comparing Philadelphia
- To nearby cities, to reflect the regional effect
- To same size cities (wrt population), to reflect the size effect

Process

Tidy Data!

Getting to Tidy Data [wickham] was a central objective of this project. Having this objective served a few purposes: (1) getting familiar with the Tidy Data concepts; (2) learning to recognize untidy data; (3) learning how to transform messy data into Tidy Data; (4) seeing how Tidy Data makes the data analysis and visualization process easier.

There are several data sets from several data sources to explore, what follows is the journey from messy data to Tidy Data.

Overview Data

Wikipedia provides some very good summary and overview data about city sizes with respect to population, seaport activity, and several other indicators [wikipedia-1, wikipedia-2]. While the Wikipedia pages are often subject to "editing by anyone," most of these data-based pages are well curated, and as such remain accurate and up to date. Getting the data contained in tables in Wikipedia pages was made rather easy by an external tool [wiki2csv].

From the Wikipedia U.S. Cities by Population page, we can begin to see where Philadelphia lies in The Big City Landscape. It is the sixth largest city in terms of population, and it has one of the busiest seaports; see Table 1.

City	2016 Population	Difference	% Difference
New York City	8,537,673	6,969,801	444.5%
Los Angeles	3,976,322	2,408,450	153.6%
Chicago	2,704,958	1,137,086	72.5%
Houston	2,303,482	735,610	46.9%
Phoenix	1,615,017	47,145	3.0%
Philadelphia	1,567,872		
San Antonio	1,492,510	-75,362	-4.8%
San Diego	1,406,630	-161,242	-10.3%
Dallas	1,317,929	-249,943	-15.9%
San Jose	1,025,350	-542,522	-34.6%

Table 1 - The Top 10 U.S. Cities by Population Size (comparisons relative to Philadelphia)

Labor Statistics Data

The Bureau of Labor Statistics (BLS) [bls] provides a comprehensive array of data related to employment, unemployment, the labor force, jobs, wages and industries with respect to the U.S. population. Several geographic perspectives into the data are available: by region, state, metropolitan area, and down to location (includes big cities and small towns). In addition, a broad spectrum of demographics related to age, gender, race and ethnicity can also be obtained. At first the massive collection of data sets is a little overwhelming, but it becomes easier to navigate once the organization and purpose of the general data set collections is understood.

The BLS provides some a good user experience and tools for navigating and selecting the data set which serves a particular interest. There are also multiple methods for getting the data: (1) direct download of

Excel, CSV, text, and PDF files; (2) a Public Data API which permits specifying exactly what data is desired, including the time frames.

The data provided by the BLS API comes in JSON format. This required implementing some parsing and extracting code. Fortunately, the BLS provides some excellent examples of using their API, and my code is modeled from their Python example.

In the accompanying Jupyter Notebook, **analysis-labor-force-population**
See the sections:
```
    -  build the data request
    -  send the request
    -  extract the JSON payload from the response
    -  extract and process the JSON payload into a PrettyTable and a plain
       Python array...
    -  create a DataFrame
```

Every BLS data set contains a SeriesID field. The SeriesID identifies the data set (employment, jobs, industries, etc.) and several additional characteristics about the data: the location, the measure. The SeriesID field violates some of the Tidy Data principles: *each variable forms a column, each observation forms a row*. The problem with SeriesID is that *multiple variables are stored in this one column*. Parsing and extracting the embedded variables from the SeriesID column was somewhat straightforward. The Python code is simple and naive, expecting that each input is for a specific type and does substring extraction based on character position and length. The BLS does provide some supporting documentation on SeriesID, but much of the knowledge was learned by navigating the BLS data portal and observing how they build a SeriesID for interactive queries.

In the accompanying Jupyter Notebook, **analysis-labor-force-population**
See the sections:
```
    -  code to: decode Months and Dates
    -  code to: decode the City/Town
    -  code to: decode the State
    -  code to: decode the Measure
    -  extract and process the JSON payload into a PrettyTable and a plain
       Python array...
```

Once the SeriesID was parsed into its discrete fields, the encoded values needed to be translated into their actual monikers. Building the lookup/translation tables for the encodings was somewhat tedious. I obtained the mappings by extracting the <OPTION> tags in HTML forms and writing some Python code to parse and build a dictionary.

In the accompanying Jupyter Notebook, **analysis-labor-force-population**

```
See the sections:
    -   code to: decode the City/Town
    -   code to: decode the State
```

Once the underlying data structure was understood, could be parsed, extracted, and transformed into something meaningful, then I was able to build a usable DataFrame of Tidy Data.

Population Data

The U.S. Census Bureau [census] data portal was simpler to interact with. The data sets were easier to identify and locate, and more readily understood. The data was downloaded in CSV format, and came in two data sets: (1) Population data from 2000-2009, (2) Population data from 2010-2016.

The Census data didn't come without its own Tidy Data problems. First, the two data sets -- while representing the exact same thing across two time periods -- were not in the same format. Column names for the same variable were different, by upper/lower case and by naming convention. This required changing and standardizing the column names in each DataFrame.

The Census data sets also had *column headers which were values not variable names*. The **melt** function was used to bring columns into rows, using the Date and Value as the existing column variables.

Finally, the two separate data sets really are *a single observational unit stored in multiple tables*. For this the simple solution was to **merge** the two DataFrames into one.

```
In the accompanying Jupyter Notebook, analysis-labor-force-population
See the sections:
    -   load the population data
```

Exploration and Analysis

Time-series

Population and employment are time-series data. A time-series is a periodic recording of an observation of the same measure over a span of time. Generally speaking, a time-series observation is made up of: (1) the name of a the measurement; (2) the value of the measurement; (3) the time at which the observation was made; and (4) the context of the measurement.

The best means to visualize a time-series is often a line chart. The time-series line chart is constructed as follows: the progression of time is represented on the x-axis, the measurement values are represented on
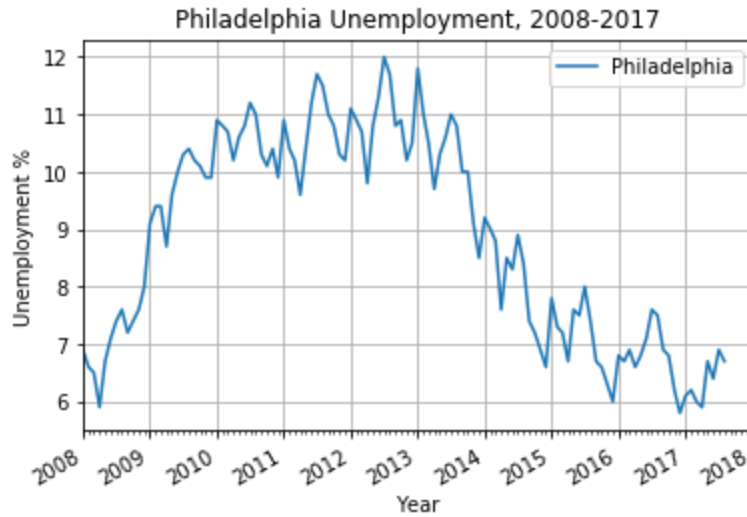
the y-axis. This type of visualization is very useful in depicting the trend of a measurement value across a span of time.

The line chart works well when displaying the U.S Population across a span of 17+ years.
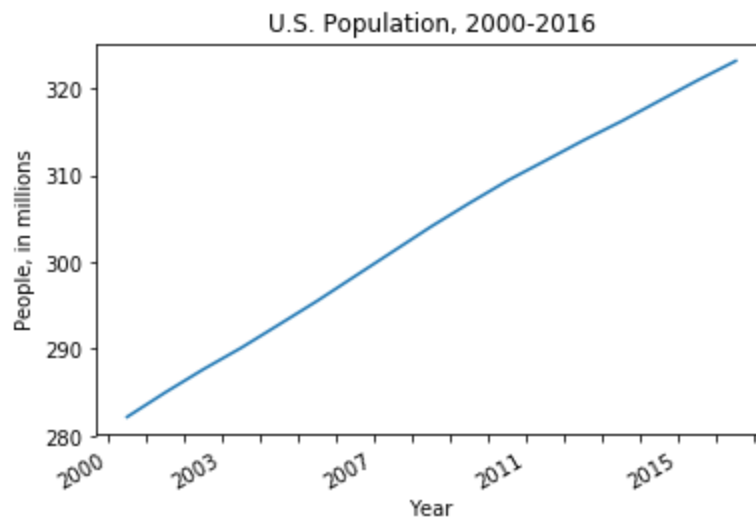

The line chart also does well when showing the unemployment rate at a national level, and in Philadelphia.

Philadelphia Unemployment, 2008-2017
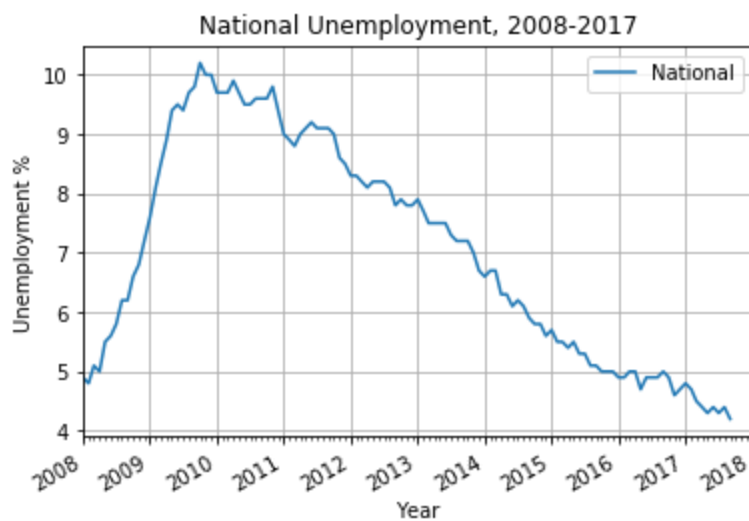
And the line chart showing the National and Philadelphia unemployment rates together provides a nice means for comparison. The Matplotlib default colors and styles also create a very visually pleasing representation of this data. Usually I don't like grid lines displayed in line charts, but in this case it adds a feel of a true economic report and I think it improves perception.

Unemployment, 2008-2017
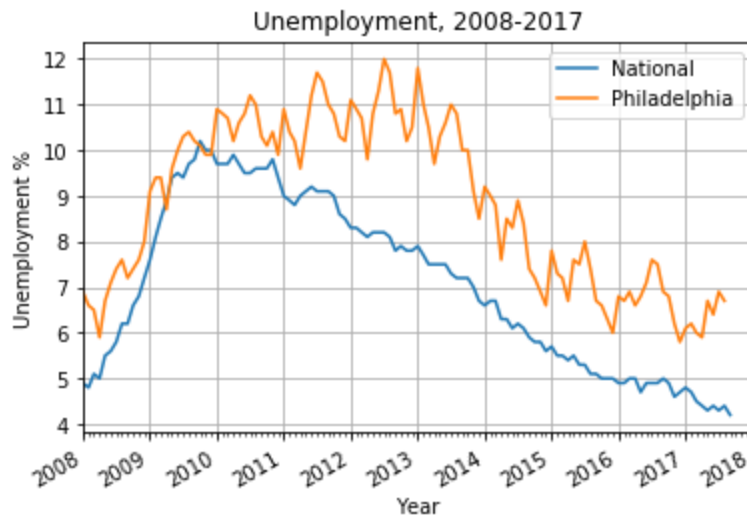
When comparing the population of the geographically close cities, we face a problem. The range in values falls into two categories: "big numbers" and "small numbers." The problem here is that New York City's population numbers are so much greater than the other cities, that we cannot see the trends very well. The solution was to plot them separately and use a shared x-axis for the date range.

Population, 2000-2016

Baltimore city

Boston city

New York city

Philadelphia city

Washington city
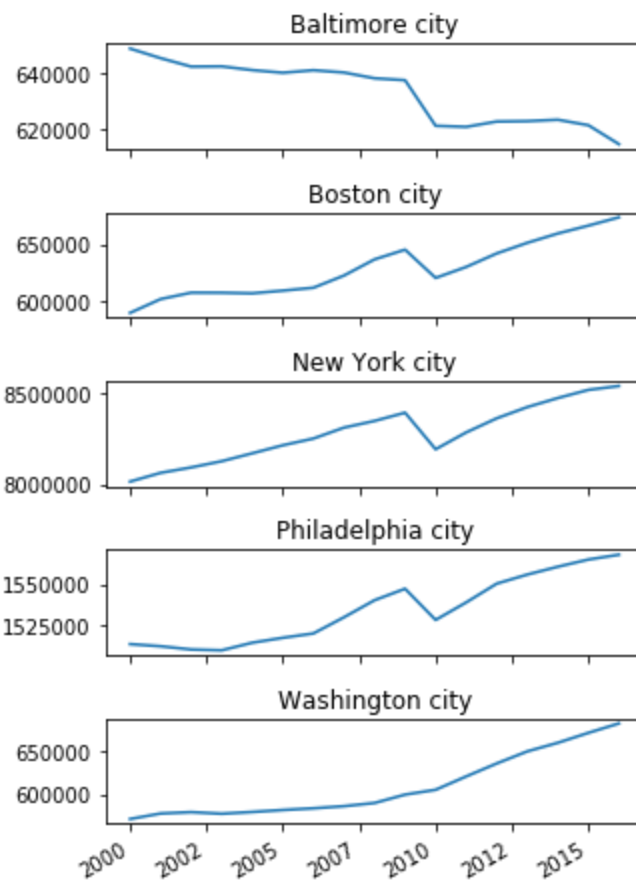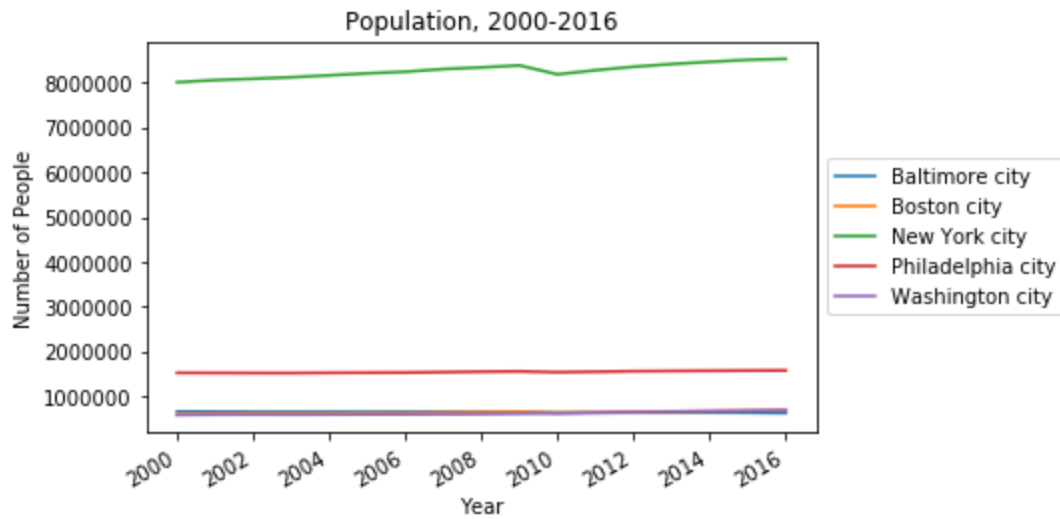
As expected, we didn't face the range disparity problem when plotting the group of five cities with the similar population sizes (because the range in population values was very close). In fact, this produced a rather nice line chart depicting the population trends. For completeness, I also created the stacked line chart with each city in separate line charts for this group. It does not have the same impression or appeal as the combined one for this group.

Population, 2000-2016
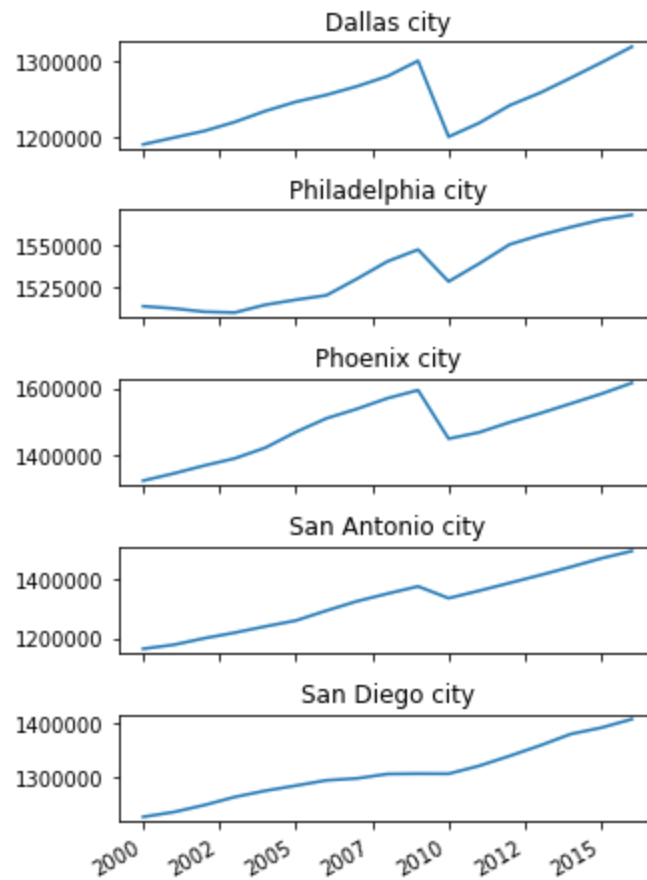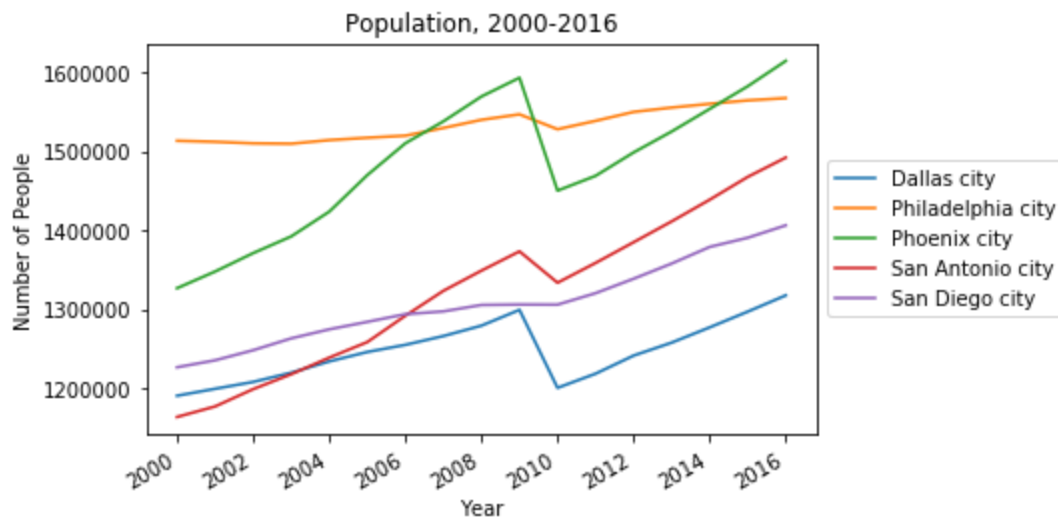
While the line chart with five city population trends produced a nice visualization, the line chart with six unemployment rates (the national rate, and five cities in a group) came out quite "messy." Comparing

unemployment rate is an important element of the analysis, so I had to find a better way to visualize this data.
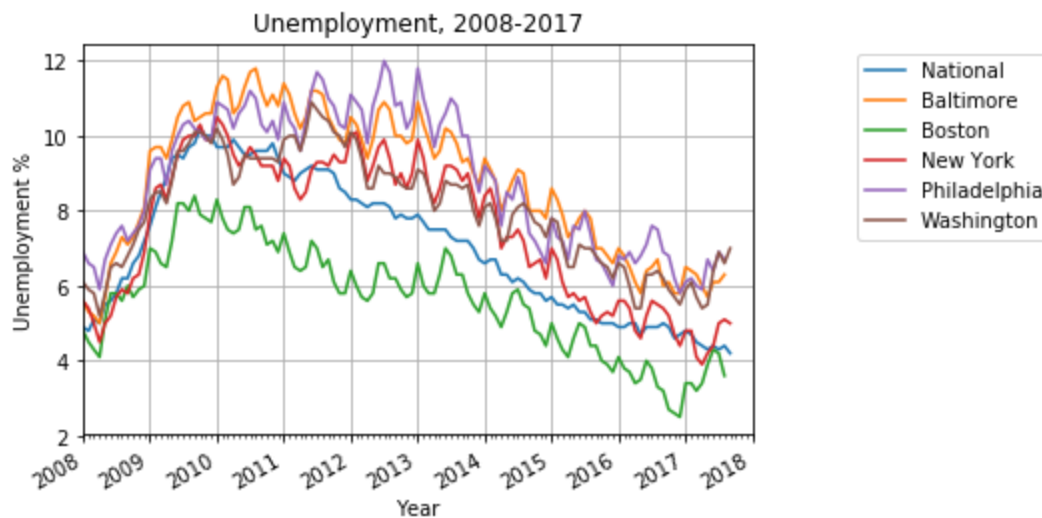
An Insightful Measure

The CCD Report introduced the measure: Change in Unemployment Rate over time. Specifically, from before The Great Recession to after (i.e., current day). They positioned this measure as an indicator of how well (or not) a city has rebounded from the recession. I thought this was an insightful measure, and decided to use this in the city comparisons. As it presents a comparison of numeric values, all within a close range, I thought a bar chart would be appropriate.
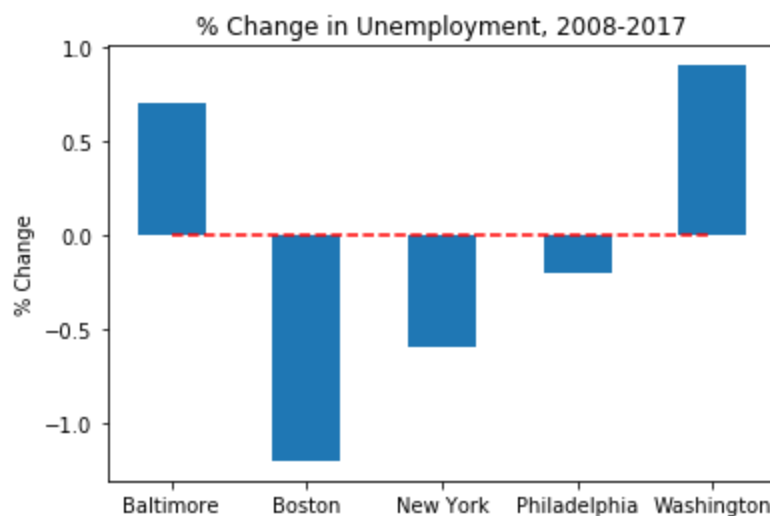

This original style of the plot (with default Matplotlib settings) has somewhat narrow bars. We will try to improve upon that in further revisions to this chart.

Results and Insights

Comparison Views

As it turns out, I think the Change in Unemployment Rate over time displays nicely as a bar chart and is very good for comparing these values across five cities. While the comparison of a delta between before and after values is nice to see, it would also be beneficial to see the trend over that time. The desire is to show the underlying unemployment rate (%) in a line chart, but we need to overcome the cluttered mess produced before. The solution is to add a separate line chart for each city, and place them using small multiples. The bar chart provides logical columns, and the wider bar width improves the visual presentation.

% Change in Unemployment, 2008-2017
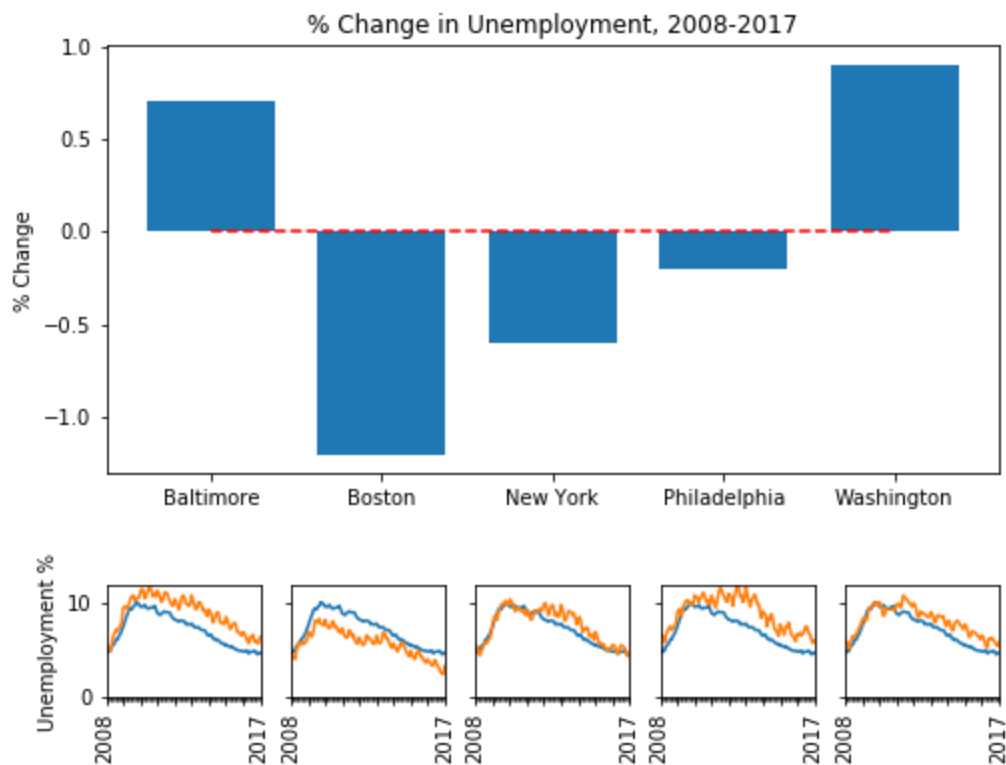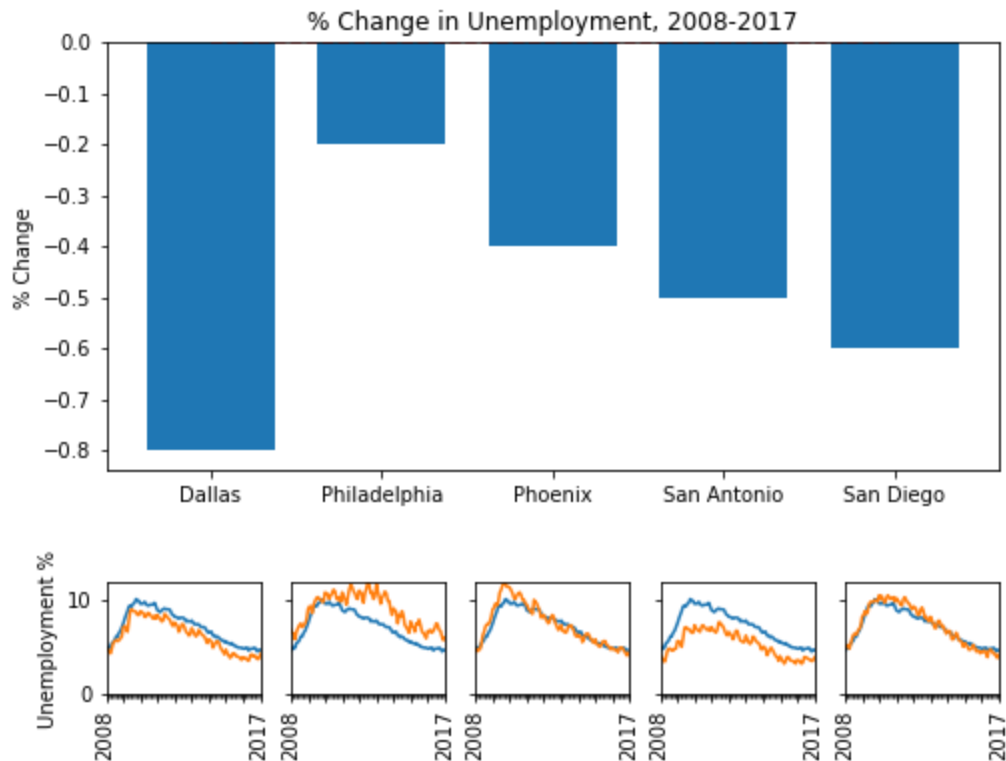
In these two combination charts we can now clearly see how Philadelphia compares with the other cities in each group. Considering the Change in Unemployment Rate over time we see: in the geo-regional group the leader would appear to be Boston, and Philadelphia is in the middle of the pack; in the population-size group Dallas is the leader and Philadelphia is clearly behind all with the lowest value.

To get some deeper perspective we can look at the accompanying Unemployment Rate line charts. Each one is plotted with the national rate (in blue). In the geo-regional group we can see that all but Boston tracked with the national rate from 2008 to 2010-2011, but then they all remain above the national rate. Boston fared very well in this time period relative to the national rate, consistently being lower.

In the population-size group Philadelphia is the only city that is notably above the national rate for almost the entire period of time. Most of the other cities in the population-size group are below or at the national rate. Overall, I would claim that San Antonio was the best performer in this group; its 0.5% improvement in Unemployment Rate over the time period is middle of the pack, but its Unemployment Rate is consistently below the national rate and better than almost every city compared (regardless of group).

Having the Unemployment Rate time-series presented in small multiples keeps the charts decluttered and easy to understand. Having the national rate in each provides two main features: (1) we can compare each city to the national rate; (2) the comparison with the national rate can be observed across charts and permits us to make a city-to-city comparison. This second feature is achieved without having an explicit chart with all of the city rates. Having the national rate in each chart gives a common point of reference

which can be carried between the individual charts. The national rate acts like an additional axis and improves cross-chart perception. While not exact, this approach does give enough precision that we can make a very good judgement of values.
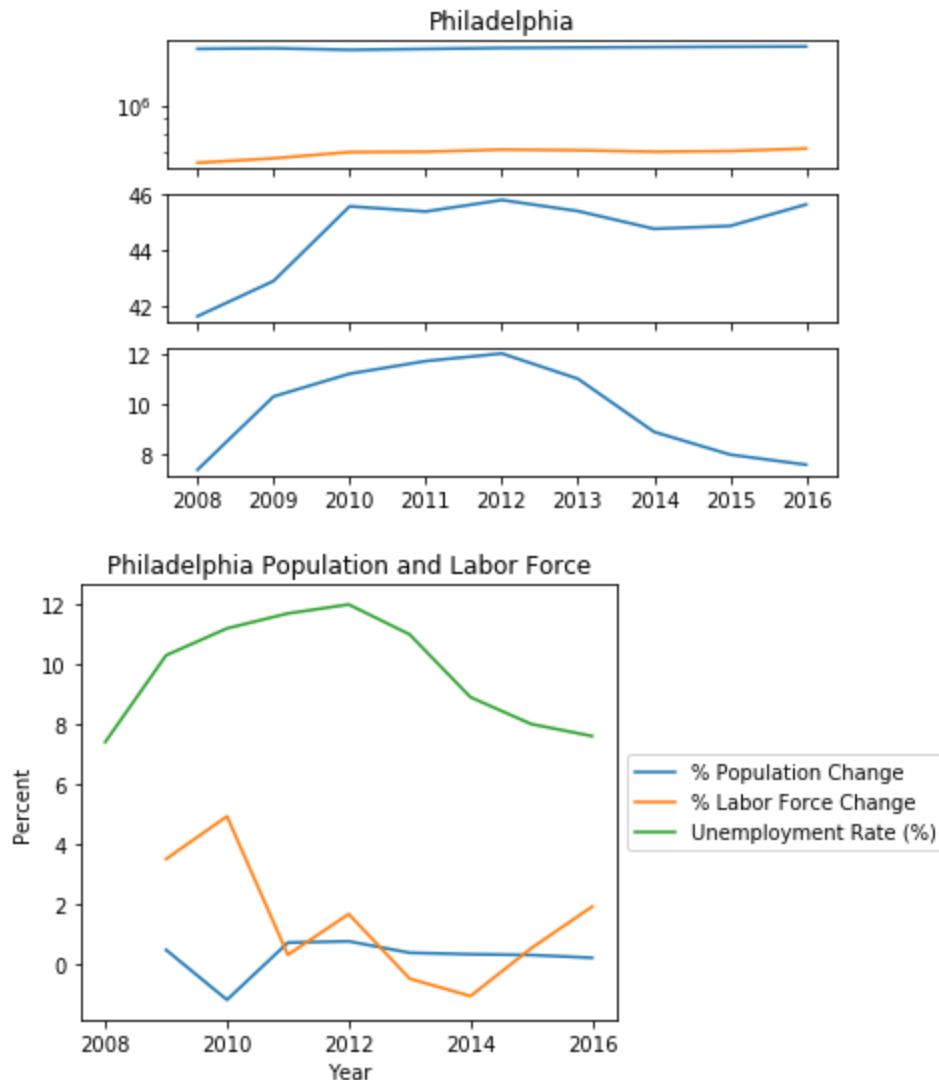
Are Population and Employment related?

The hypothesis that population and employment are related, and could be predictive of each other was derived from some commentary in the CCD Report.

Note that in general urban population saw a decline (a "blip") in the time of the Great Recession. People moved out of the cities, the population numbers declined. That must have had some impact on the unemployment rates? Plotting the ratio of labor force[1] to population should show some sort of trend. The line charts in most of the cities show a rise and then flattening out. This would suggest that the decline in population made up for the amount of jobs lost due to the recession.

---
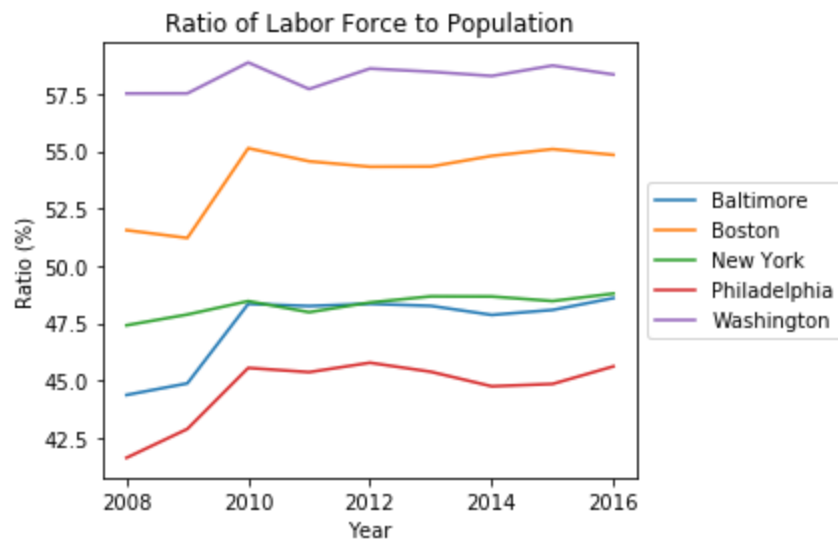
[1] The Bureau of Labor Statistics defines *labor force* as: "the Labor Force includes all persons classified as employed or unemployed in accordance with the [BLS] definitions.

Philadelphia

Philadelphia Population and Labor Force

The plots are just confusing, and don't reveal anything. Perhaps the hypothesis is misguided. Population and Employment may be related, but are not necessarily predictive of each other.

A deeper read of The CCD Report's commentary provides some insight into who was moving out of the cities. Their conclusion was that young families were moving out because there was a decline in the quality of public schools and programs. This was due to the economic downturn (fewer jobs, less tax revenue). The "flee" probably exacerbated the issue because generally it was the middle class citizens that were leaving the city. This left the city with unbalanced ends of the socioeconomic spectrum: the upper class, and the lower class. Jobs and employment became scarce for those in need, and one of the major supporting revenue sources was shrinking. Cities had a difficult time coping and mostly responded with tax increases and program cuts (education is hit with teacher layoffs and school closings). State and federal funding for cities also declined (because of the economic downturn, and political spite) and that contributed to the deeply dramatic effect on public school funding.

This whole thing turned out to be a much bigger social and economic issue, which the city is trying to insulate itself from such dramatic effects of in the future. How? By diversifying its image and expanding into growing industries. This is done through gradual tax and government policy reform, and enticing new businesses to start or move here.

So this "population-employment hypothesis" did not pan out. What should we explore next because of these findings? We should look at other dimensions of data to see if there is an indicator which can be drawn out and help answer the question: *is Philadelphia going to have a "big boom" in its future?*

This is the main point which sets up Part II.

Future Work

Coming in Part II

The Change in Unemployment Rate over time with Unemployment Rate trend visualization creates a basic framework for assessment and comparison of cities. This approach will be used to establish the "Gold Standard" cities: those with the best numbers and trends, and perhaps sprinkle in less tangible qualities like general perception of being a "booming city."

We will look at the Change in Unemployment over time across a broader spectrum of cities to see which are performing the best (establishing the Top 5 or 10, and the best in various geographic regions). An important data dimension will be added: city employment breakdown by industry, and by occupation.

With these standard bearers, we can evaluate and assess where Philadelphia lies and see whether its trends are beginning to follow the established leaders. From this we can begin to draw some conclusions, and answer the nagging question…

As an ultimate deliverable, I aim to build a visualization tool which will permit a user to choose a city and see how it fares in this model. And with some very general heuristics predict whether that city is about to boom or bust.

References

[bls]   https://www.bls.gov/data/

[brookings]   https://www.brookings.edu/topic/cities-regions/

[ccd]   http://www.centercityphila.org/research-reports/philadelphia-an-incomplete-revival

[curbed-1]   https://philly.curbed.com/maps/philly-tower-highrise-construction-map

[curbed-2]

                  https://philly.curbed.com/2017/8/9/16059724/philadelphia-megadevelopment-megaproje
                  ct-list-information

[shrm]

                  https://www.shrm.org/hr-today/trends-and-forecasting/labor-market-and-economic-data/
                  Documents/philadelphia.pdf

[usa-data]   https://www.data.gov

[usa-stats]   https://www.usa.gov/statistics

[wickham]   https://www.jstatsoft.org/article/view/v059i10

[wikipedia-1]   https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

[wikipedia-2]   https://en.wikipedia.org/wiki/List_of_ports_in_the_United_States

[wiki2csv]   http://wikitable2csv.ggor.de

Additional Material

(not cited, but worth reading for additional background)

[citizen]   http://thephiladelphiacitizen.org/paul-levy-economic-growth-philadelphia/

[inquirer]   http://www.philly.com/philly/business
                 http://www.philly.com/philly/business/20151206_The_region_s_2016_economic_outlook.html

[newsworks]   http://www.newsworks.org

[npr-twoway]
                 http://www.npr.org/sections/thetwo-way/2017/09/12/550447620/census-bureau-median-incomes-rose-and-poverty-levels-fell-in-2016

[opendata]   https://www.opendataphilly.org

[philabusj]
                 https://www.bizjournals.com/philadelphia/datacenter/interactive-map-socioeconomic-breakdown-of-every.html

[philadata]   http://www.phila.gov/slash-data/

[phillymag]   http://www.phillymag.com/tag/economy/
                 http://www.phillymag.com/citified/2015/03/02/philadelphia-economic-growth-among-lowest-worldwide/
                 http://www.phillymag.com/articles/philadelphia-new-boom/
                 http://www.phillymag.com/articles/philadelphia-real-estate-boom/
                 http://www.phillymag.com/business/2016/11/21/philly-real-estate-forecast-2017-no-irrational-exuberance-here-not-yet/
                 http://www.phillymag.com/citified/2016/03/24/philly-metro-area-7th-largest-in-country/

[state-impact]   https://stateimpact.npr.org/pennsylvania/