**Literature Survey**

Previous research using construction data as basis for analysis, show that such data can yield meaningful insights, in terms of trends and linkages to events [SB-1][SB-5].

We seek to build upon earlier efforts. Some earlier efforts used outdated technology (e.g., ESRI ArcGIS) and outdated methods (e.g., Microsoft Office Excel and Access) to organize data [SB-4]. Other efforts used effective data analytics techniques, but deficient visualizations [SB-2]. We can improve visualization by replacing static diagrams with interactive, dynamic and better practices (avoid red-green color schemes. [SB-2].

We consider various techniques for analyzing [SB-2] [SB-3] [SB-6] our baseline building construction data set.

We consider - join with complementary data, based on news articles, which we could harvest either by web-scraping or directly using API. Either way, candidate data set would be a corpus of unstructured data. Plan would be to use techniques such as TD-IDF [SB-7], or perhaps more cutting edge methods [SB-8] to analyze the data.

**Appendix - Reference**

[SB-1] Clustered housing cycles

Summary**:** Analyzes city 'building permits' over time, and argues that despite the strong linkages often found between housing market cycles and business cycles at the national level, the relationship between housing and the business cycle is less clear with subnational data because housing markets are highly localized. [SB-1]

[SB-2] Housing and the City: A Spatial Analysis of Residential Building Activity and the Socio-Demographic Background in a Mediterranean City, 1990–2017

Summary: Analyzes changes over time in building activity relevance of 12 spatially-explicit indicators derived from the analysis of building permits, over a 25+ year time period. Uses hierarchical clustering analysis. While findings may be useful, we argue that visualization can be dramatically improved, using other visualization techniques, such as replacing static diagrams with interactive d3, and not using graphics that rely on red-green as color scheme. [SB-2]

[SB-3] A Spatial Analysis of Red X Properties and Correlation to Foreclosed Properties

Summary: Focus is drawing correlation between economic policy and location for buildings in city of Chicago. Less relevant, but we can borrow ideas on how to convey analysis using maps. [SB-3]

[SB-4] Tracking neighborhood development and behavioral trends with building permits in Austin, Texas

Summary: Interesting research, but dated technology: ESRI’s ArcGIS ArcMap was the primary software tool used in this research. Additionally, Microsoft Office Excel and Access programs were used to organize and sort data. [SB-4]

[SB-5] Crisis-driven changes in construction patterns: evidence from building permits in a Mediterranean city

Summary: Analysis of localized construction activity, over time, and correlation to crisis events. Looks analysis of spatio-temporal changes in building permit data as a promising tool. It can be used from both a research and policy perspective to interpret the long-term expansion of large cities, and linkages to economic crisis. Paper argues that by understanding complex metropolitan systems’ dynamics increasingly requires methodologies integrating exploratory data analysis with (non-parametric) inferential approaches.

[SB-6] Fast Proximity Graph Generation with Spark

Summary: Depending on the type of analysis and data set size, this could be a technique to analyze the data.

[SB-7] Hot Topic Detection Based on a Refined TF-IDF Algorithm

[SB-8] Semantic Unsupervised Automatic Keyphrases Extraction by Integrating Word Embedding with Clustering Methods