*Executive Summary*

The American healthcare system has undergone some of the most sweeping, industry-wide regulatory changes over the last eight years with the introduction of the Affordable Care Act in 2010. One of the most influential elements of the law is the MLR (medical loss ratio), which requires insurers to spend 80% of the premiums they collect no matter the outcome of the actual services rendered. Simple logic dictates that to grow the 20%, insurers must spend more of that 80% of a fully-insured market’s premiums. While providers are privy to this, they continue to bill more and more to generate more revenue. This has created a positive feedback loop that usually leaves the employer and their employees financially “in the dust.”

*Motivation*

My motivation to pursue this topic primarily stims from direct experiences dealing with the frustrations seen at my previous position at a health insurance brokerage along with my interest in cost-containment solutions to mitigate these problems for employers that I will be investigating in my new position. Because of the dramatic rate increases being seen in the small business, fully-insured market across the country, reference-based pricing models and self-funding (self-insured) methods are become more financially advantageous for larger swaths of the employer market, demanding more price transparency and accountability beyond the veil of secrecy draped over the health insurance industry.

*Data Question*

At its root, my motivation to explore this topic comes from the desire to explore the misaligned incentives of the modern, American healthcare system. More specifically, the cost discrepancy for procedures over the course of time. I would like to do most of my analysis, exploration, and preparation in Python to then be able to use this code in an R Shiny application.

1. *Can I quantify the price inflation for inpatient procedures from 2011-2015?*
2. *How do price discrepancies between providers within similar markets highlight the lack of transparency currently experienced by all parties that interact with the healthcare system?*
   1. *Medicare (stable, fair value market price) vs negotiated rates between providers and insurance companies on the private market*
3. *Could I predict the average cost of a procedure for 2016 using regression techniques using Sklearn?*

Articles and broader research that have sparked my thinking:

1. <https://www.huffingtonpost.com/2013/05/08/hospital-prices-cost-differences_n_3232678.html>
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4693848/>
3. <http://content.time.com/time/subscriber/article/0,33009,2136864,00.html>
4. <https://www.economist.com/news/leaders/21736138-welcome-doctor-you-revolution-health-care-coming>

*Schedule (5/26 – 7/24)*

1. Proposal Due (5/24/2018)
2. Get the Data (done)
3. Clean & Explore the Data in Python (6/14 – 6/21)
4. Prepare data for Shiny in Python(6/22 – 7/7)
5. Build & Deploy Shiny App in R (7/8 – 7/19)
6. Demo Day (7/24/2018)

*Data Sources*

1. The primary source of my data will come from the CMS (<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient.html)>, comprising of national Medicare data in a variety of diagnostic categories from 2011 – 2015.
2. CMS also hosts their own database (<https://data.cms.gov>).