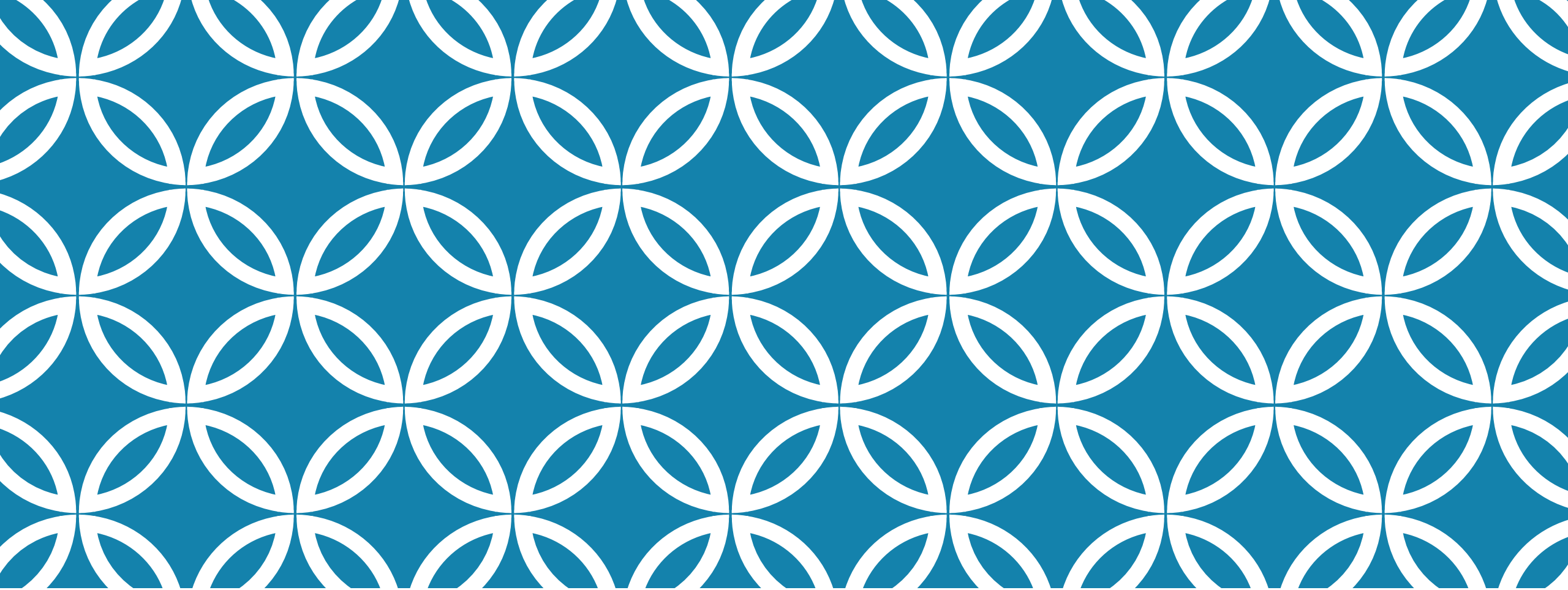

# CAPSTONE PROPOSALS

Dana Hackel
DSIR 720

# DRUG COSTS

I will use Wholesale Acquisition cost and National Average Drug Acquisition cost data to build a Multiple Linear Regression model that predicts drug price values to be used by pharmaceutical or insurance companies.

Some trends to investigate:

- What is correlated with drug cost?
- Hypothesis: specific molecular entities will increase drug cost if they cost more to manufacture.
  - Might be hard to gather data – drug class
- For each year the drug exists, how much does the price decrease?

Data will be collected from:

- https://healthdata.gov/dataset/prescription-drug-wholesale-acquisition-cost-wac-increases
- https://healthdata.gov/dataset/nadac-national-average-drug-acquisition-cost
- https://healthdata.gov/dataset/nadac-comparison
- https://healthdata.gov/dataset/prescription-drugs-introduced-market
- Drugs@FDA API

My MVP (Minimum Viable Project) is: a LR model and trend analysis

My stretch goals include:

- Build a Fire Dashboard
- Clustering to improve trend analysis

My observations will be drug name (trade and generic), entities, MoA, MDR (dispense), class, dosage, and package size and my target will be price prediction.

# CANCER

I will use  Cancer Registry data along with CDC Wonder, SEER, and HINTS data to build two Classification models that predicts  Cancer Type to be used by doctors when diagnosing patients.

Trends to investigate

- Type of cancer by age, race, sex, location
- Type of drugs used for treatment
- Where can prevention tests be implemented?
  - For a type or a location

Data Sources:

- https://healthdata.gov/dataset/cdc-wonder-cancer-statistics
- https://healthdata.gov/dataset/cancer-incidence-surveillance-epidemiology-and-end-results-seer-registries-limited-use-0
- https://healthdata.gov/dataset/health-information-national-trends-survey-hints

MVP – classification models and trend analysis

Stretch Goal

- Recommender System – for cancer drug which has best outcome based on predicted type of cancer
- Also a Fire Dashboard?

My observations will be patient age, sex, race, location and my target will be cancer type.

# SUGGEST DRUGS FOR PATIENTS 65 AND OLDER

I will use Pillbox, Drugs@FDA, and Prescriber Checkup Data (for patients 65 and older is free), to gather information about over the prescribed drugs to create a classification model to predict type of drug needed.

Things to investigate

- Comorbidities
- Drug class used
- ??

- https://pillbox.nlm.nih.gov/developers.html
- drugs@FDA (using FDA's API for all drug label information)
- https://www.propublica.org/datastore/dataset/prescriber-checkup ($200 for student access)
- https://www.propublica.org/datastore/dataset/medicare-part-d-prescribing-data-2012-seniors (Free)

MVP – classification models and trend analysis

Stretch Goal

- Recommender System

My observations will be symptoms and my target will be drug to mask symptoms.

# OTHER POSSIBLE RESOURCES DEPENDING ON WHERE MY PROJECT GOES

https://healthdata.gov/dataset/prescription-drugs-introduced-market

https://healthdata.gov/dataset/unintentional-prescription-drug-deaths-0

https://healthdata.gov/dataset/nchs-drug-poisoning-mortality-state-united-states-1

https://healthdata.gov/dataset/1990-through-2011-national-survey-drug-use-and-health-nsduh

https://healthdata.gov/dataset/1992-through-2010-treatment-episode-data-set-admissions-teds

https://healthdata.gov/dataset/vital-statistics-opioid-related-deaths-gender-and-raceethnicity-beginning-2003

https://healthdata.gov/dataset/vital-statistics-opioid-related-deaths-age-group-beginning-2003