Predicting and Characterizing High Cost Type II Diabetes Patients

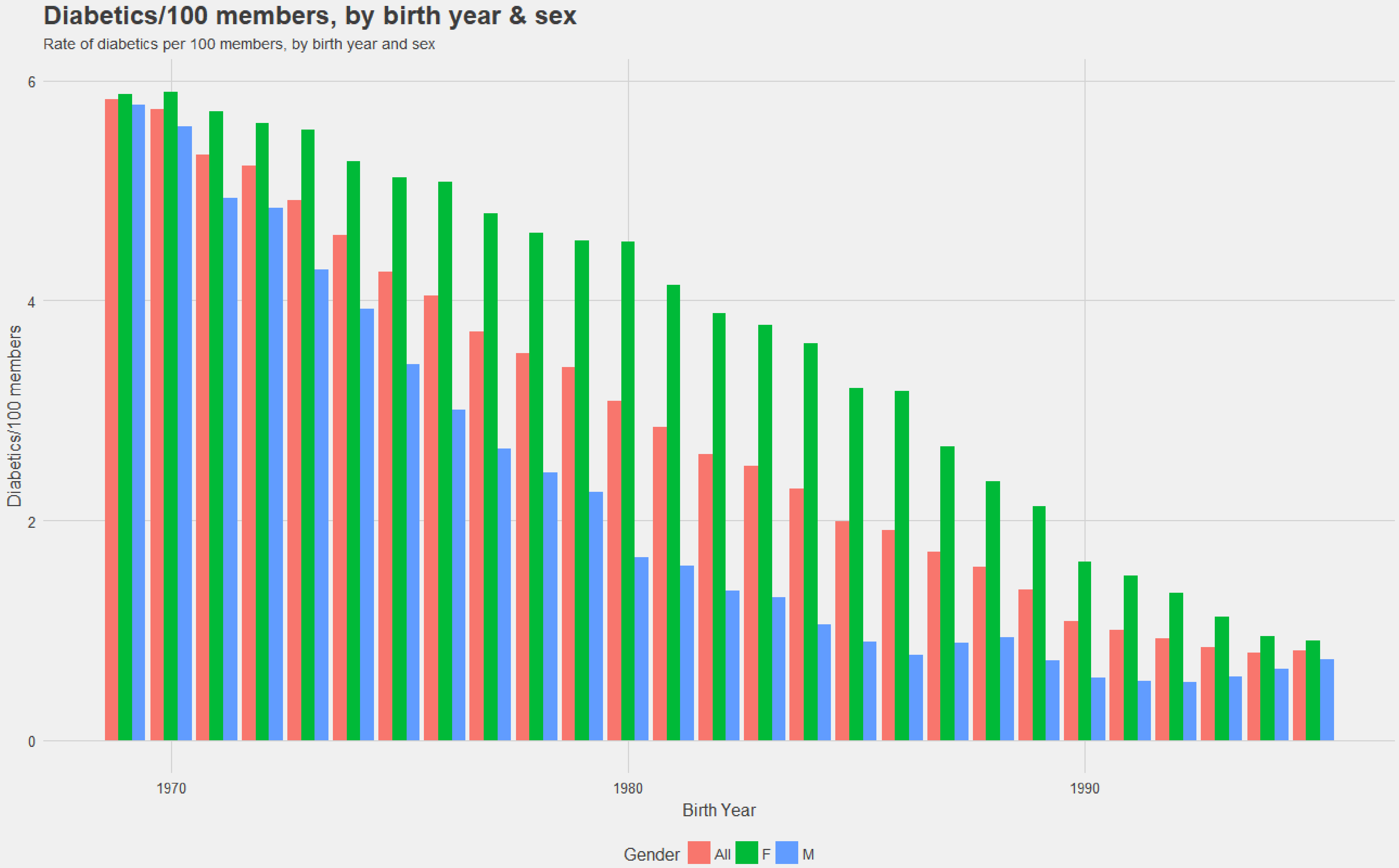
MOHAMMED MODAN (THE SHU-MEN)

MACALESTER COLLEGE

MMODAN@MACALESTER.EDU

The Data

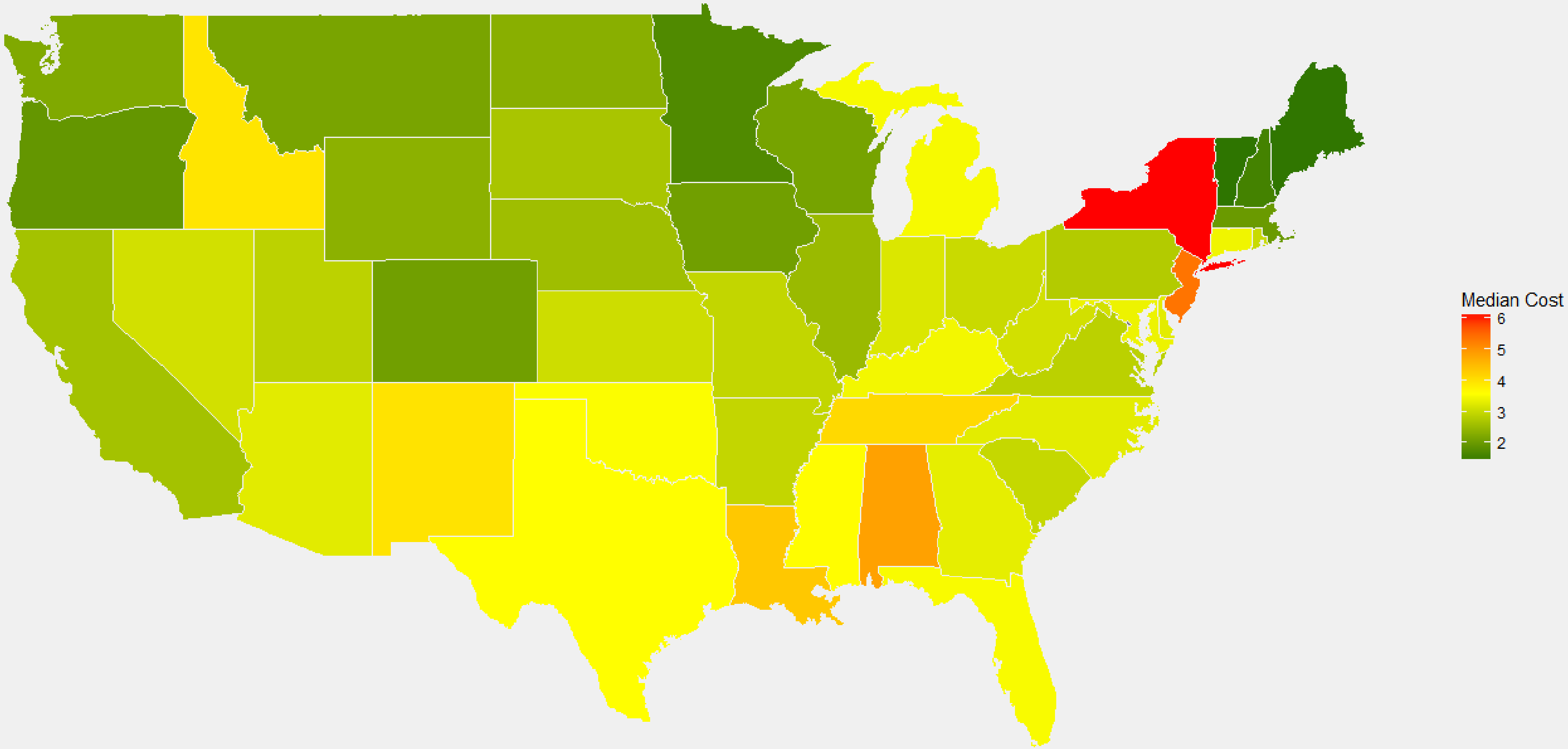
- Insurance claims data from 39133 type II diabetics
- Medical, pharmacy, confinement claims + lab results
- Variables include provider data, diagnosis data, drug class, etc.

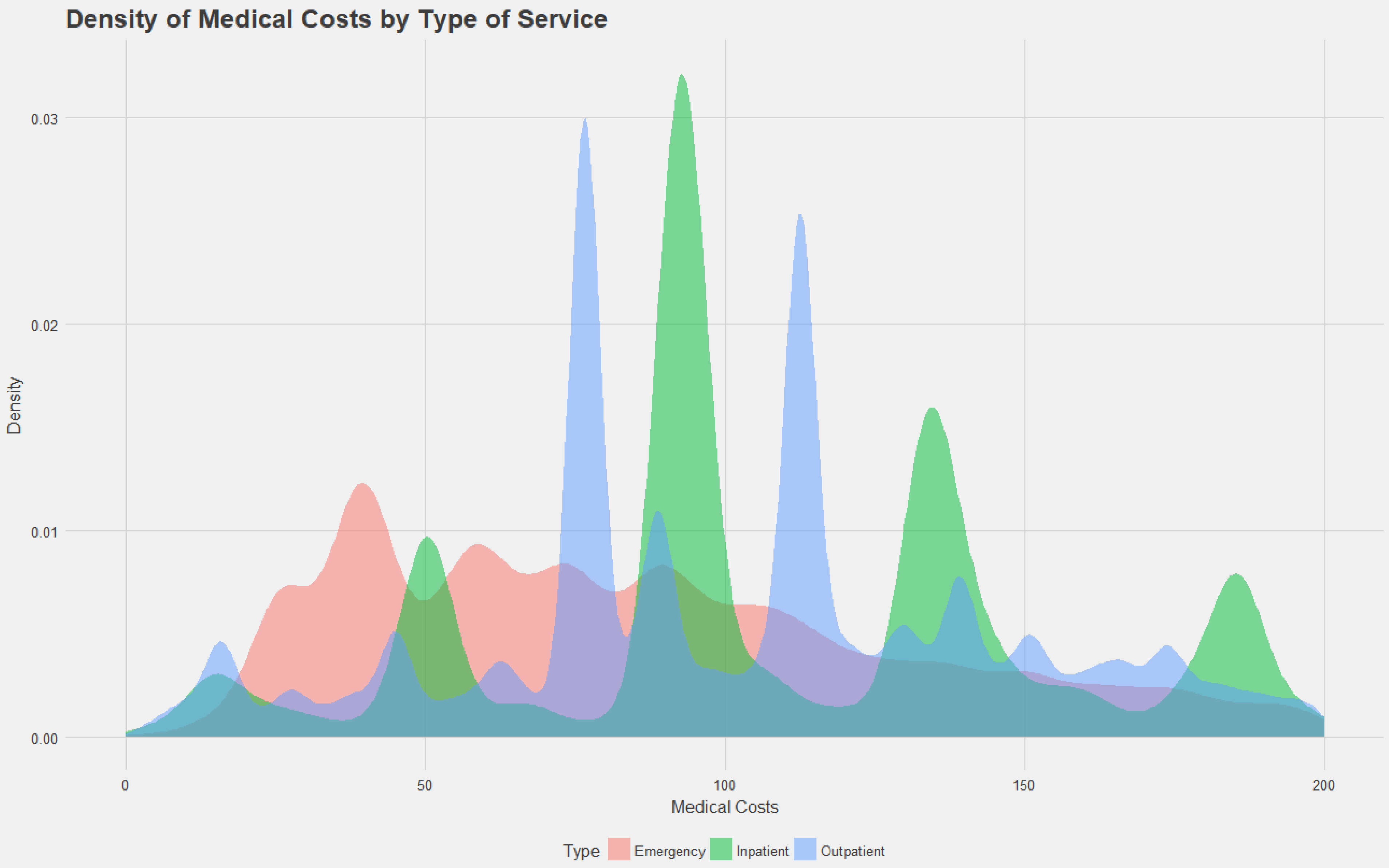


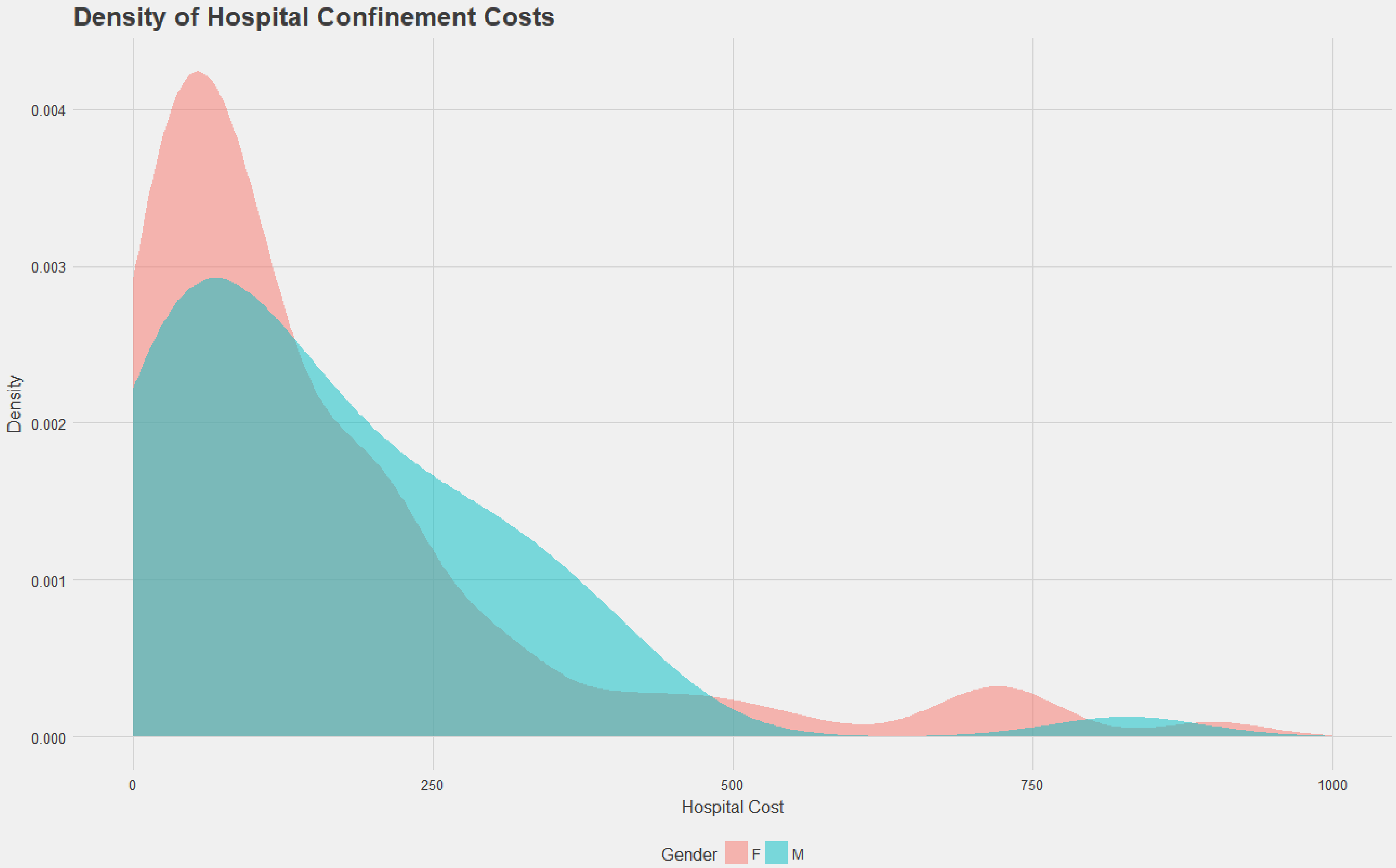
Diabetics/100 members, by birth year & sex

Rate of diabetics per 100 members, by birth year and sex

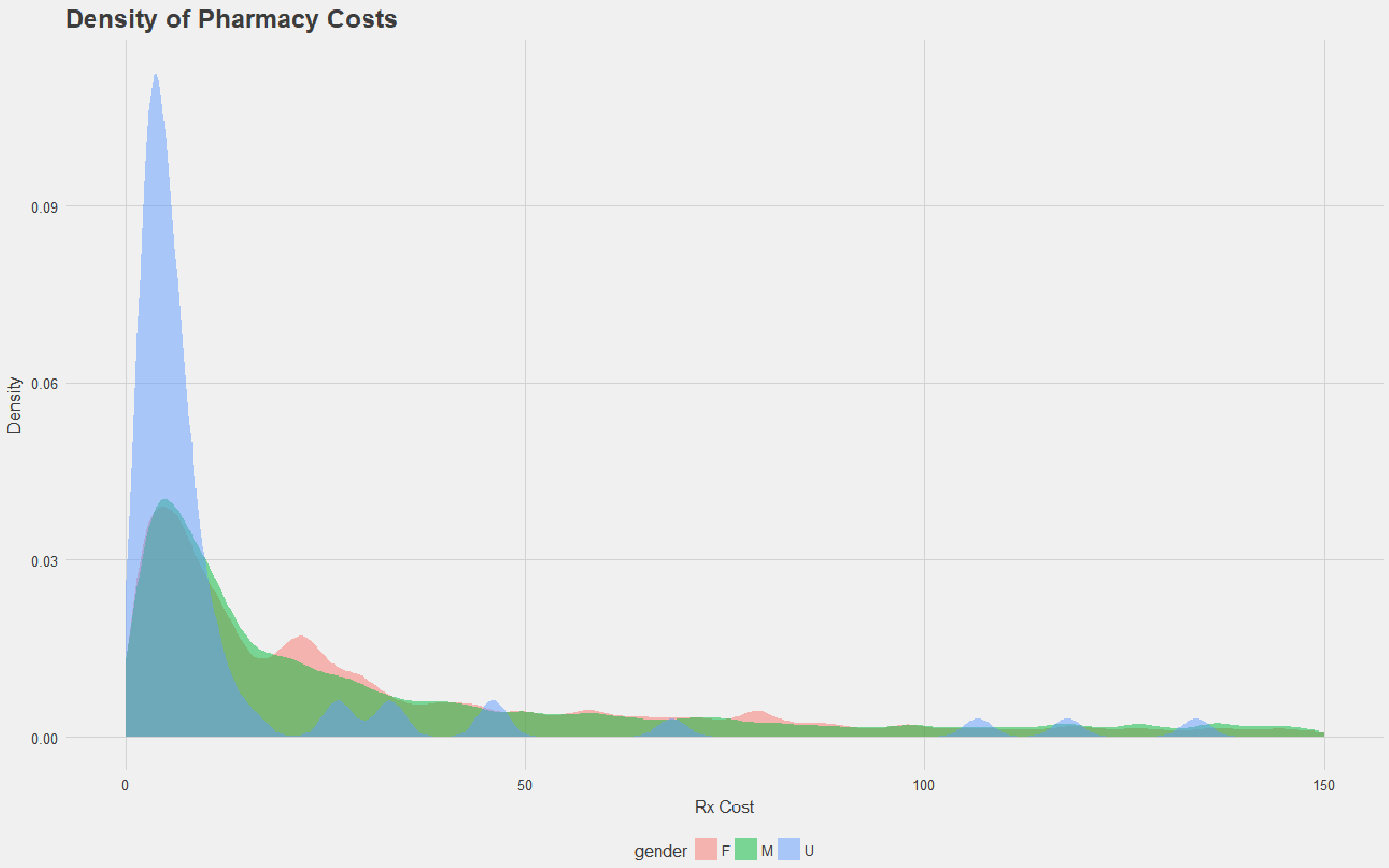
# Diabetics per 100 members by state




Density of Medical Costs by Type of Service
Density
0.03
0.02
0.01
0.00
0
50
100
150
200
Medical Costs
Type
Emergency
Inpatient
Outpatient



Density of Hospital Confinement Costs


Density of Pharmacy Costs
Density
0.09
0.06
0.03
0.00
0
50
100
150
Rx Cost
gender
F
M
U

The Task - Prediction

- Task – predict 6000 highest cost patients
  - Numeric prediction or classification

- Strategy
  - Predict costs, take top 6000

    OR

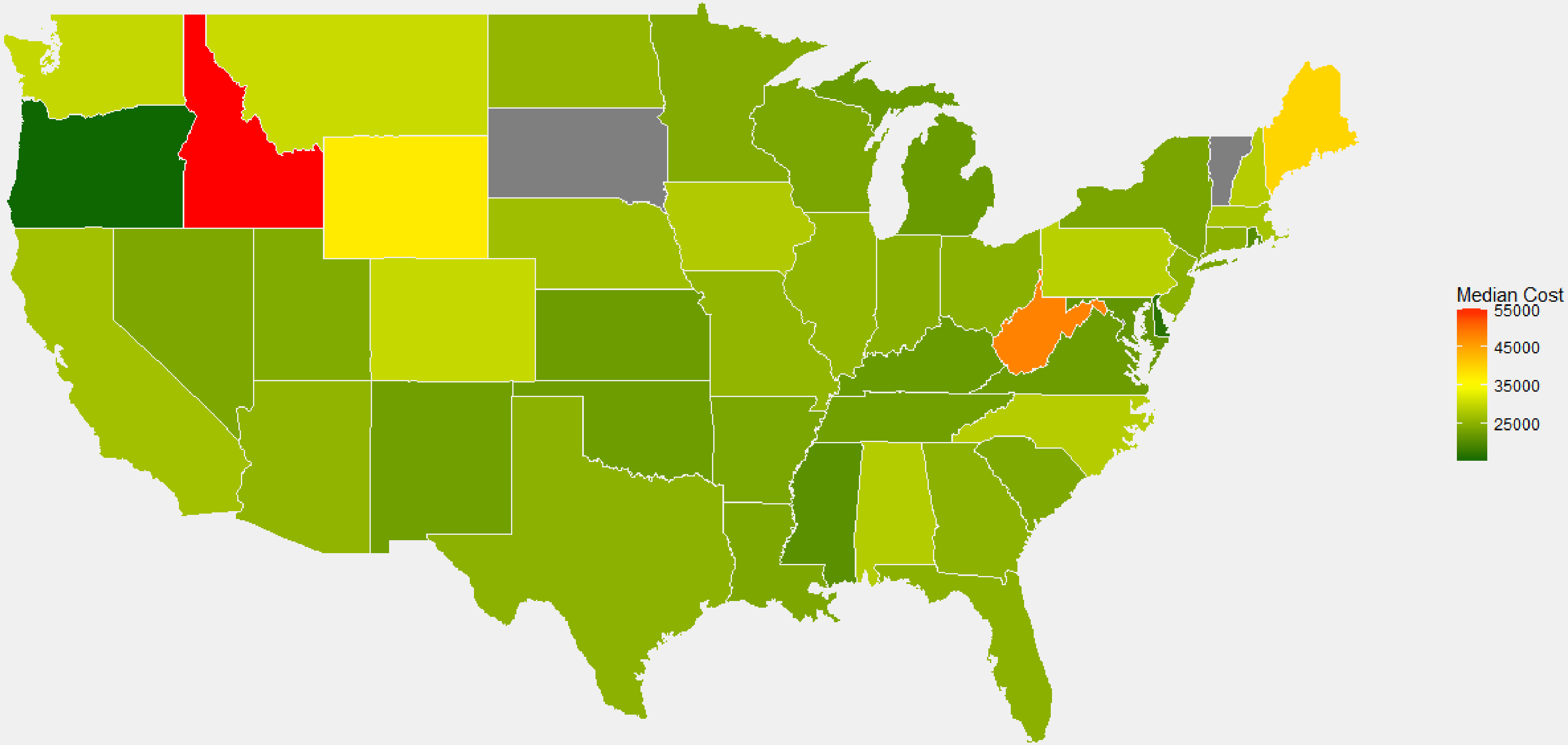
  - Take top 6000 most probably classified as top 6000

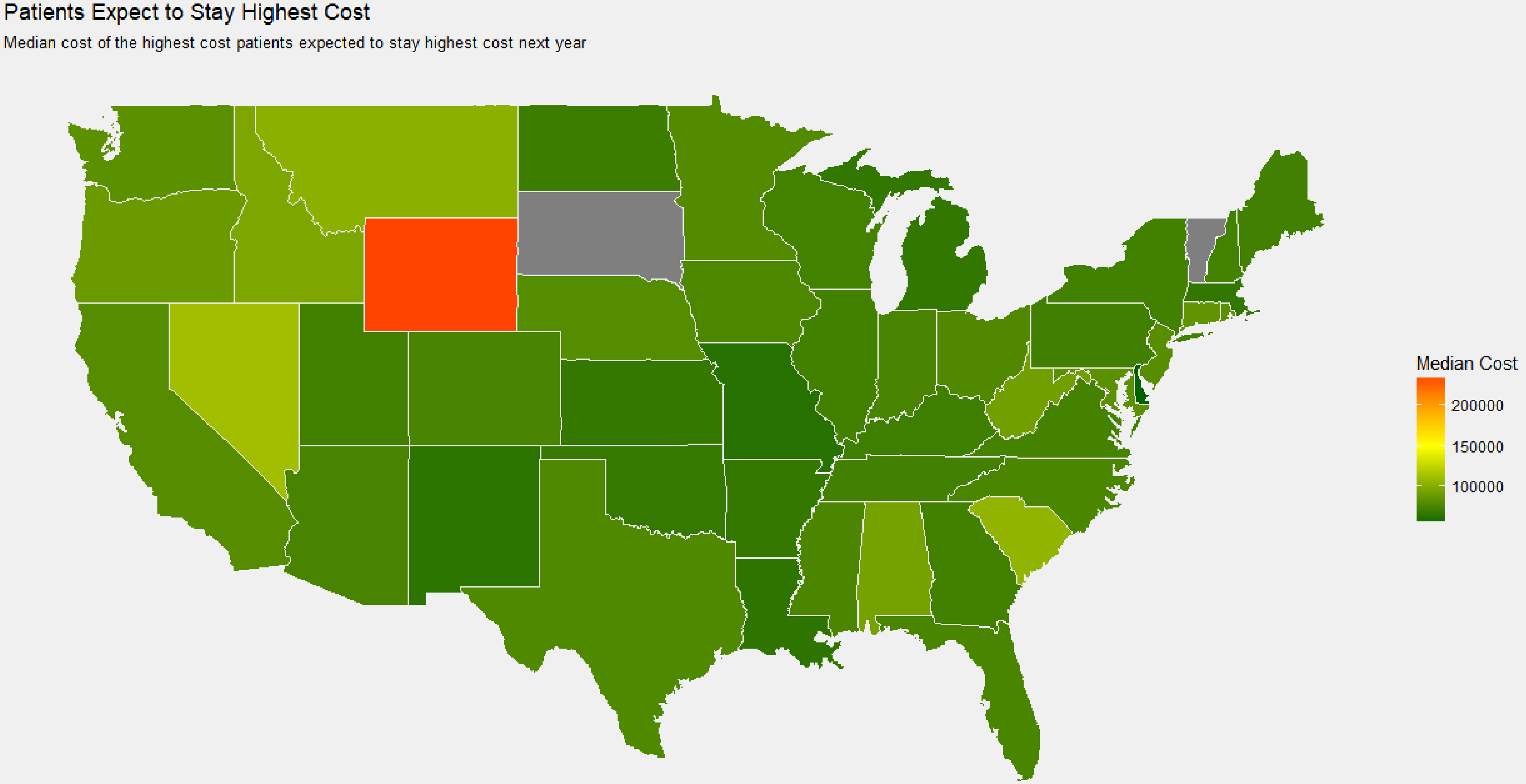
Getting Ready

1. Clean data (feasibly) in training and target
2. Remove trauma and pregnancy claims
3. Join data tables by patient and days from diagnosis
4. Attempt various summary statistics
   - Collapse table to one entry per patient
5. Visualize trends and model vs target data

6000 Expected Highest Cost Patients

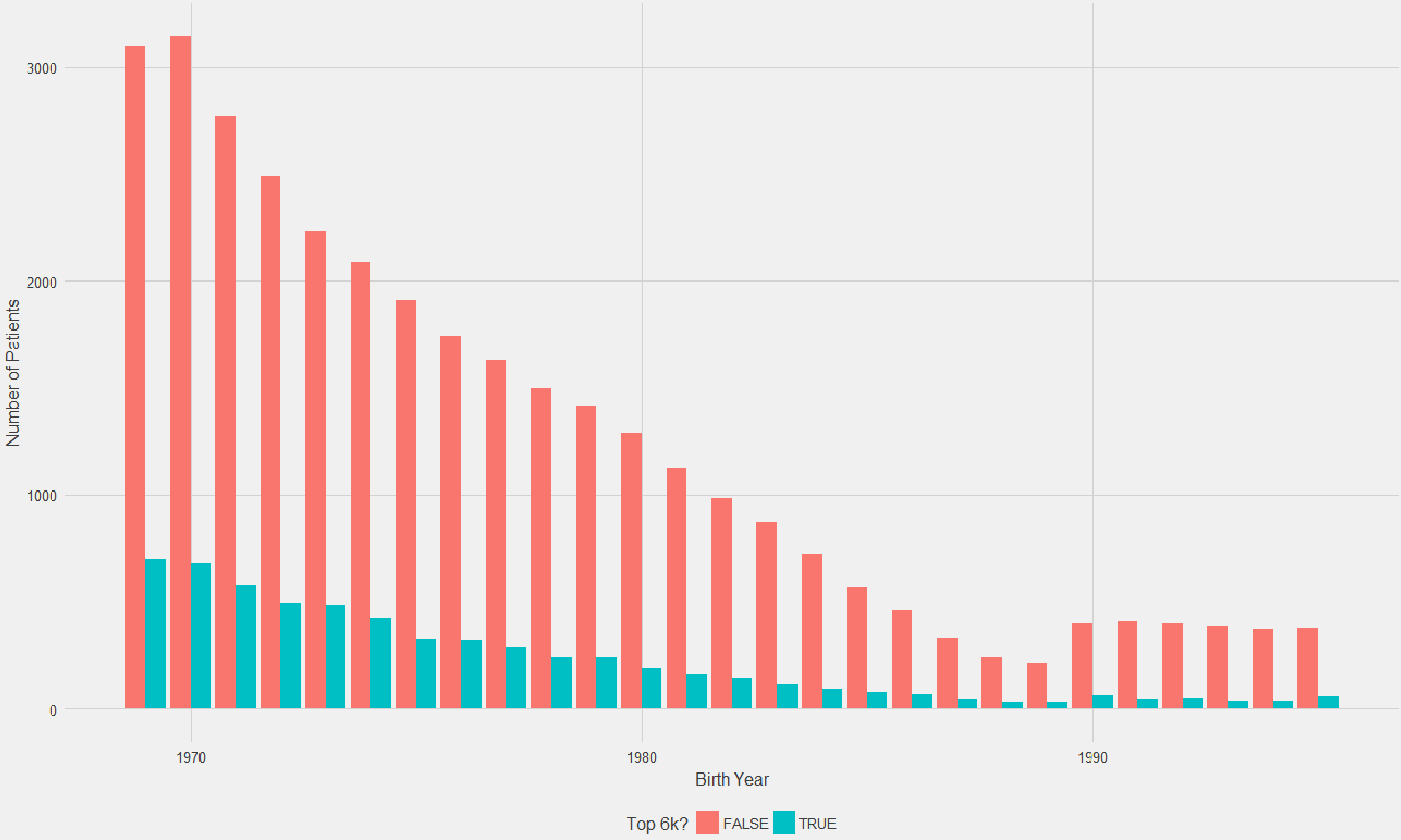
Median cost of 6000 expected highest cost patients by state

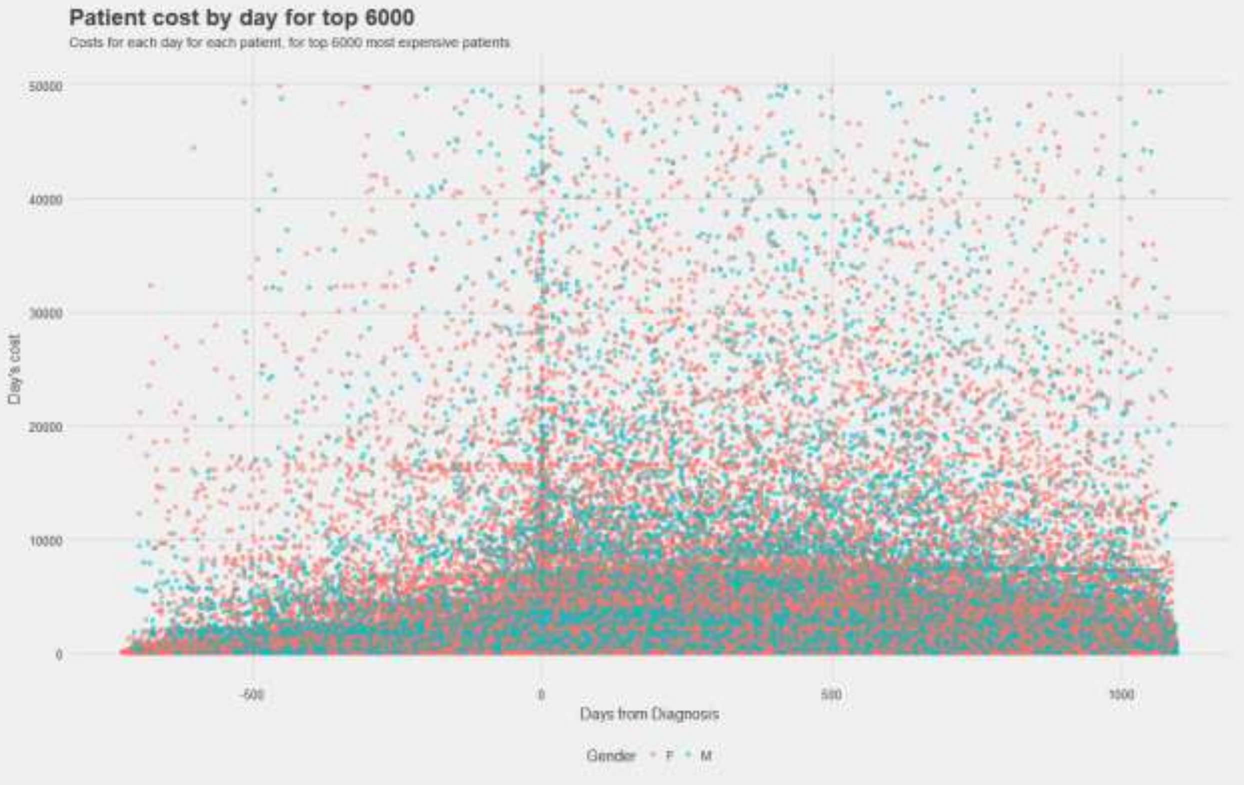




Patients Expect to Stay Highest Cost
Median cost of the highest cost patients expected to stay highest cost next year
Median Cost
200000
150000
100000

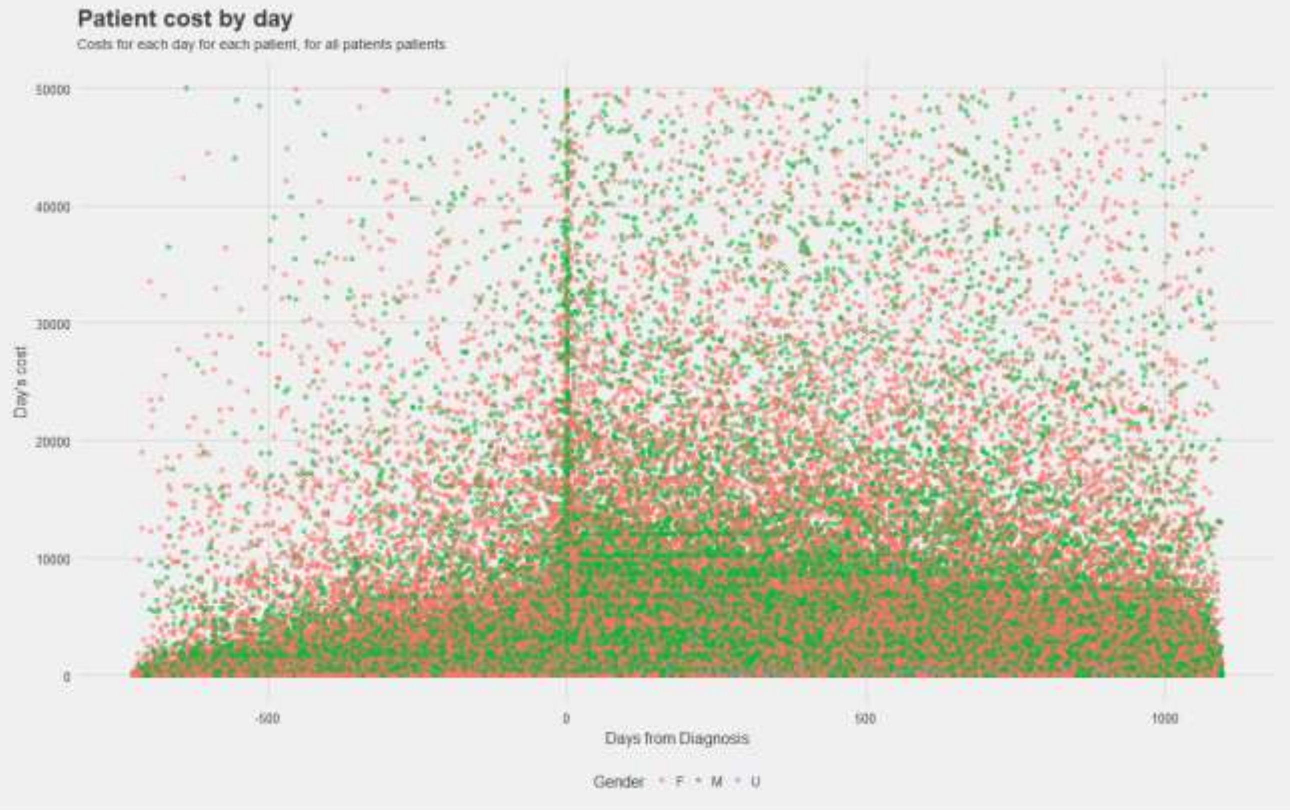
Age distribution of top 6k vs others





Patient cost by day for top 6000

Costs for each day for each patient, for top 6000 most expensive patients

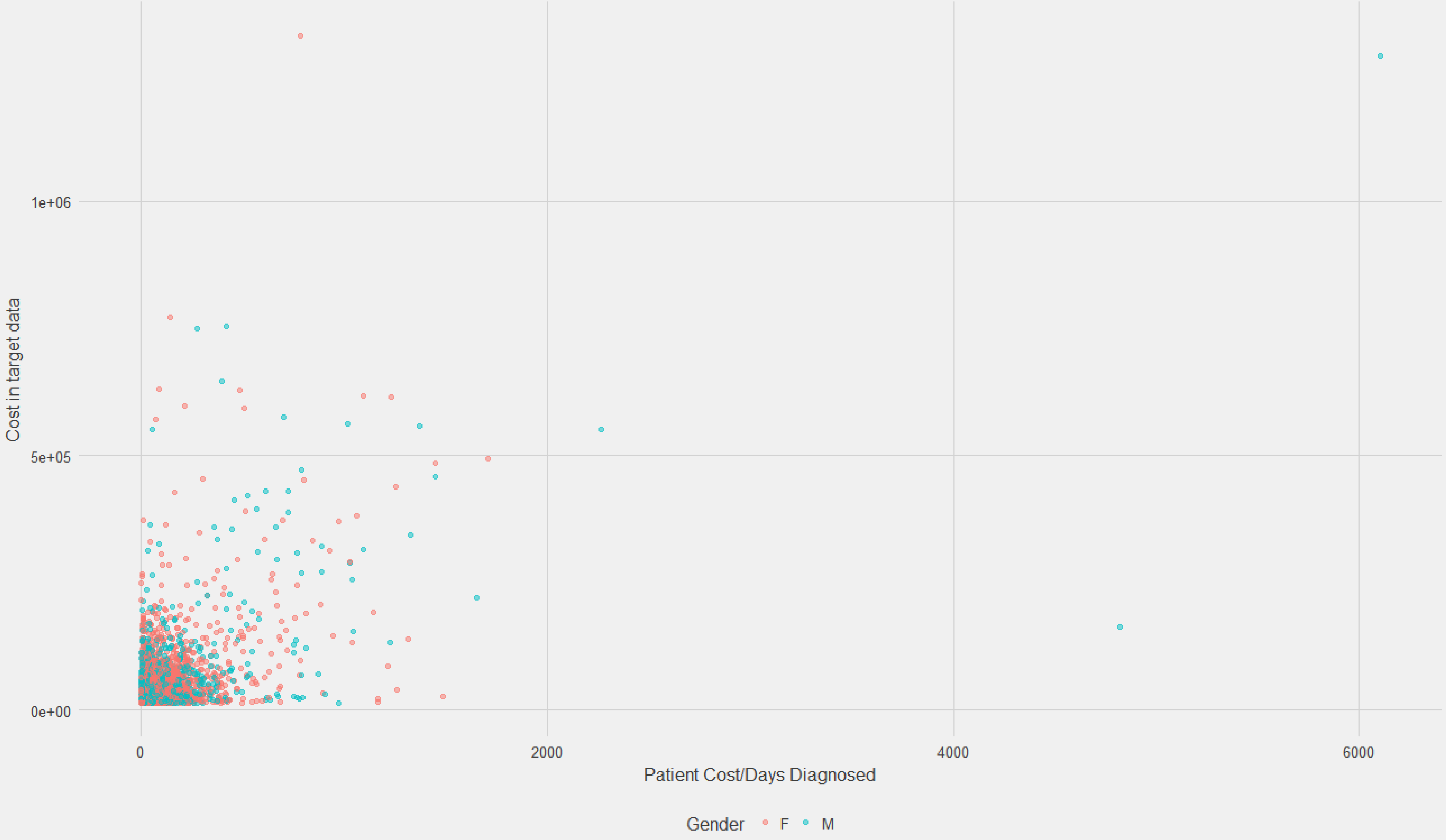


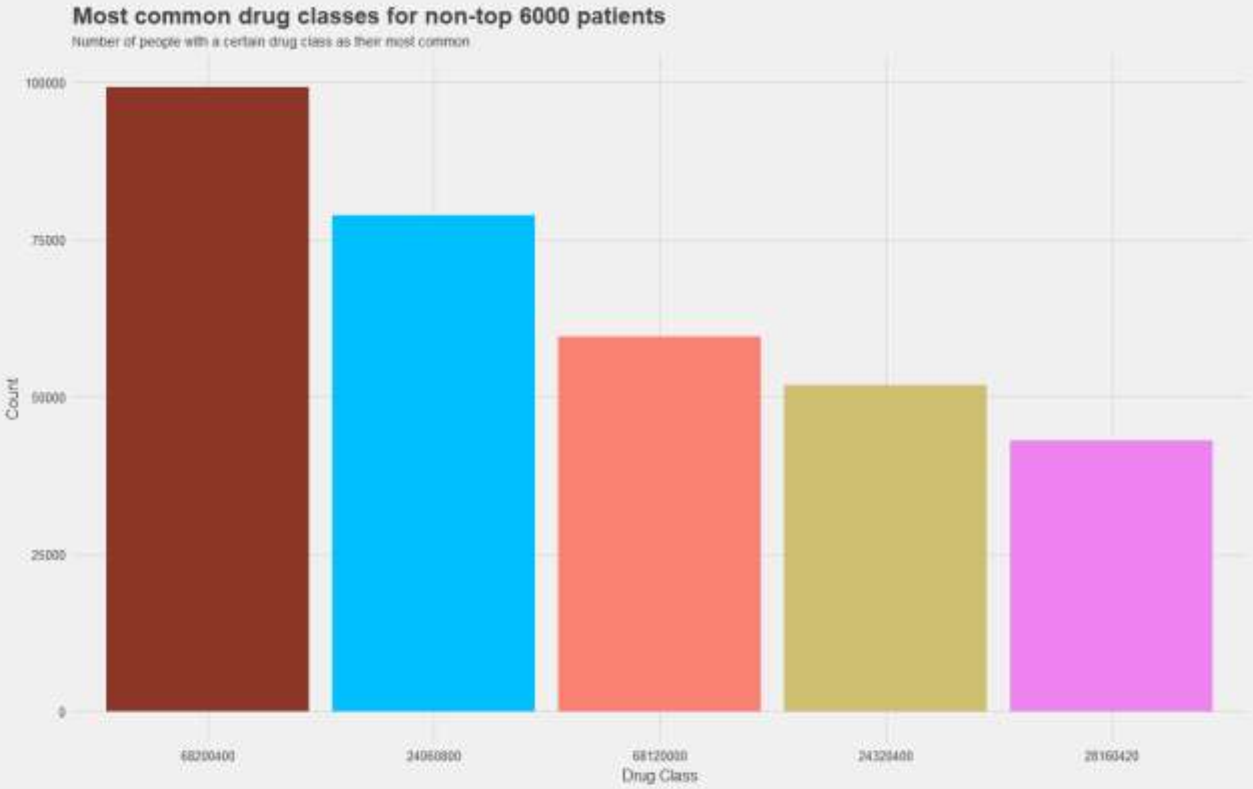
Patient cost by day

Costs for each day for each patient, for all patients patients

Target Cost vs Training Cost

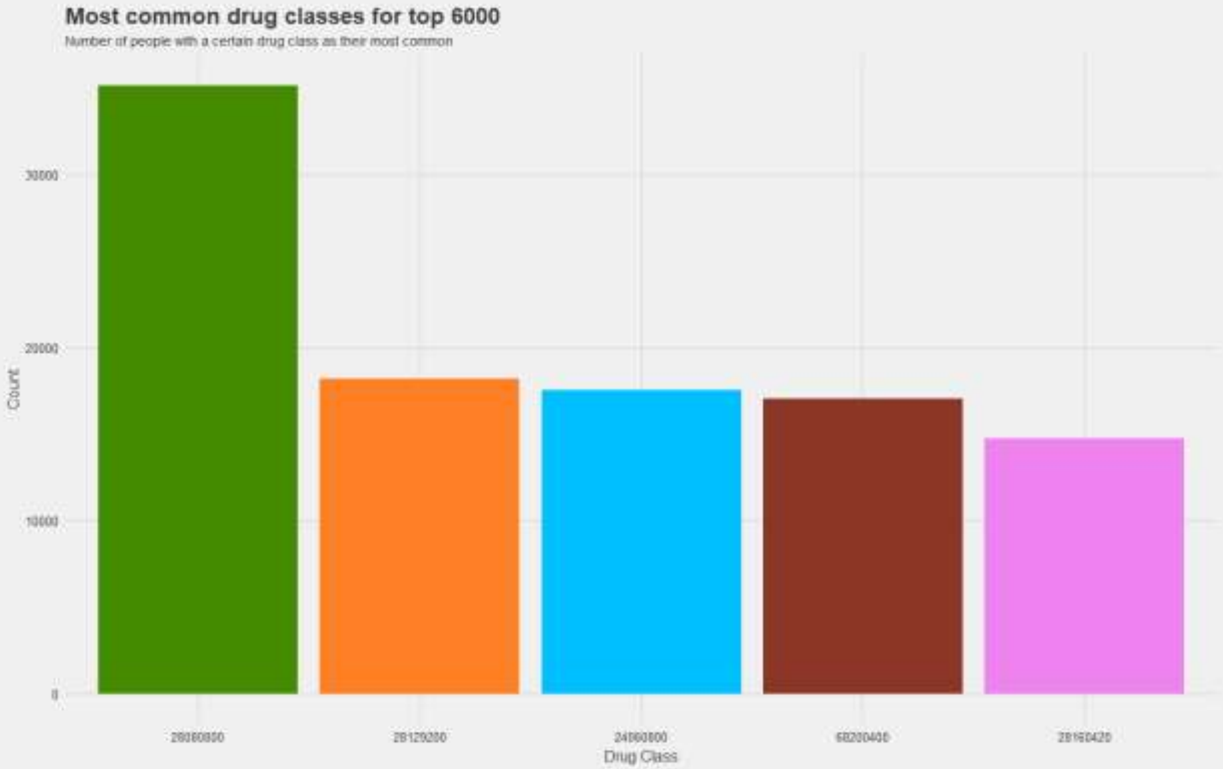
Total cost in training/days diagnosed to predict target cost, minus trauma/pregnancy





Most common drug classes for non-top 6000 patients

Number of people with a certain drug class as their most common



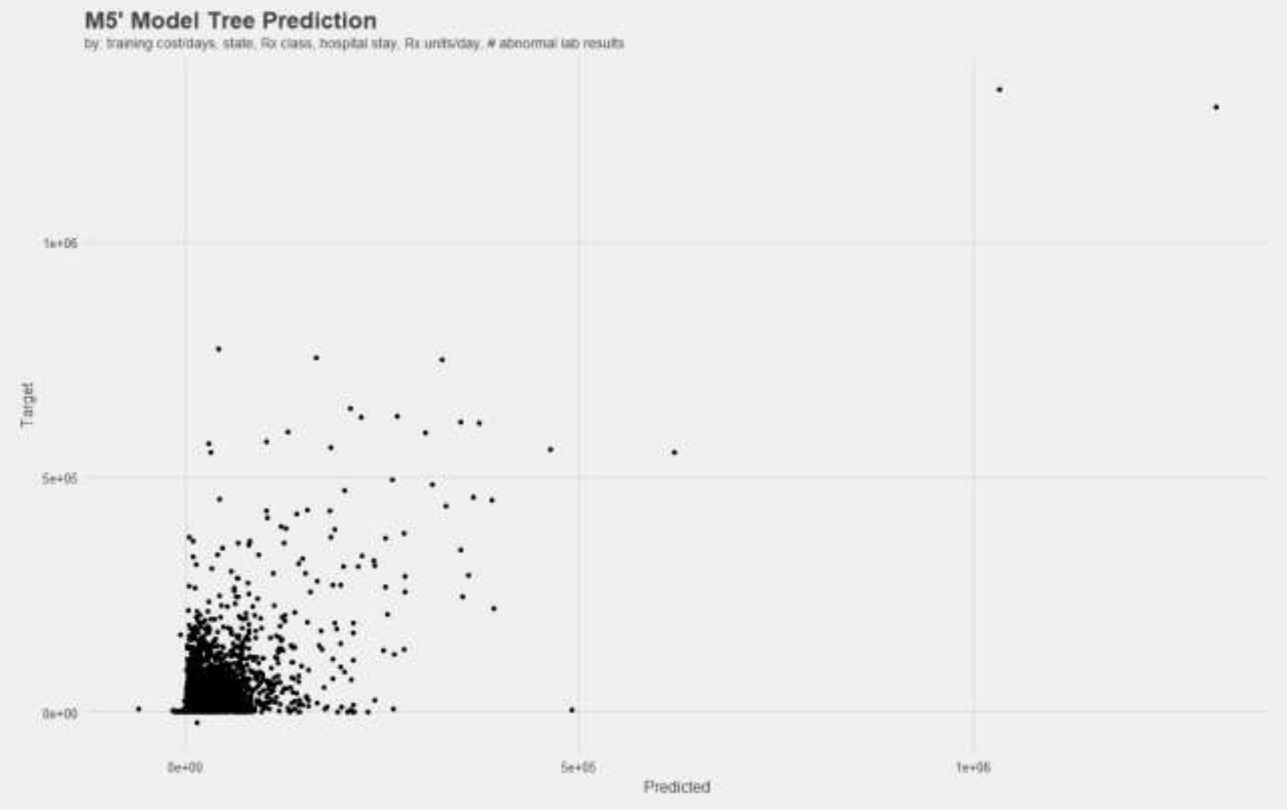
Most common drug classes for top 6000

Number of people with a certain drug class as their most common

Predictive Model

- M5’ model tree from:
    - Cost/days diagnosed
    - State
    - Most common Rx class
    - Avg Rx units per day
    - Avg hospital stay
    - Normalized # abnormal lab results

- 53% correct in test, 69% correlation, $7435 MAE



Looking to the future

- Not the best model, limited by variables explored
  - Future: explore more summary statistics

- Explore further as a classification problem
  - Most variables categorical, limited in regressions

- Teamwork makes a difference
  - Team commitment can allow for a distribution of the workload