Build a Predictive Model for Heart Disease

By: Edward Kerr

Table of Contents

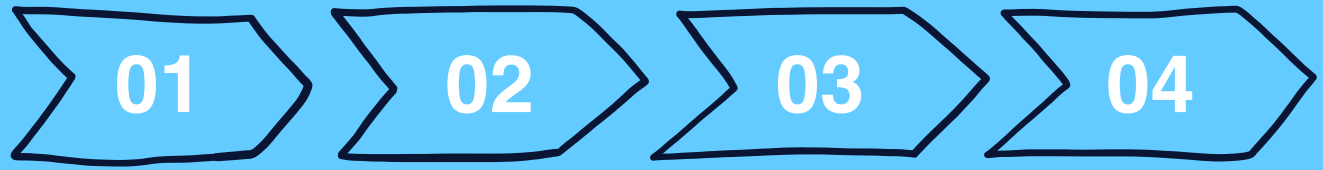
Workflow

Get Data

Kaggle data set and various related websites

Charts

Created charts on Google Sheets and Tableau



EDA

Cleaned and organize data on Google Sheets and Tableau

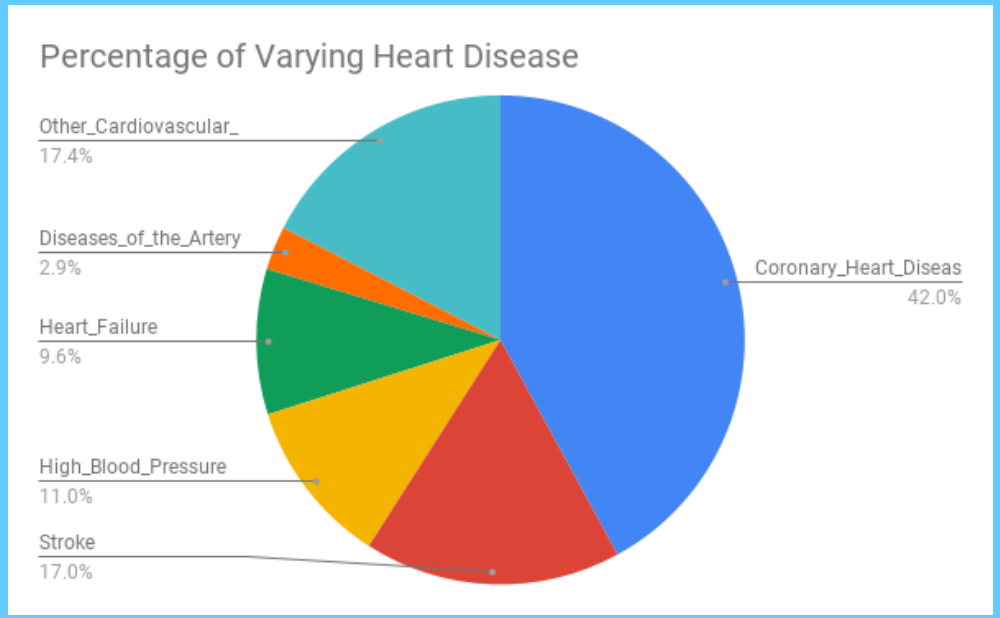
Analyze Charts

Analyze Charts

Introduction

- Heart Disease is the leading cause for death at 659,000 people in the United States.

- Goals are to recommend building a data science model to predict heart disease and lower hospital cost.



$363 Billion

- From 2016 to 2017.
- Amount spent annually by hospitals in US.
- Includes cost of health care services, medicines, and lost of productivity due to death.

Methodology (Pathology)

Data set

Data set was collect from Kaggle

For EDA

Using Google sheets and Tableau.

Google Sheets

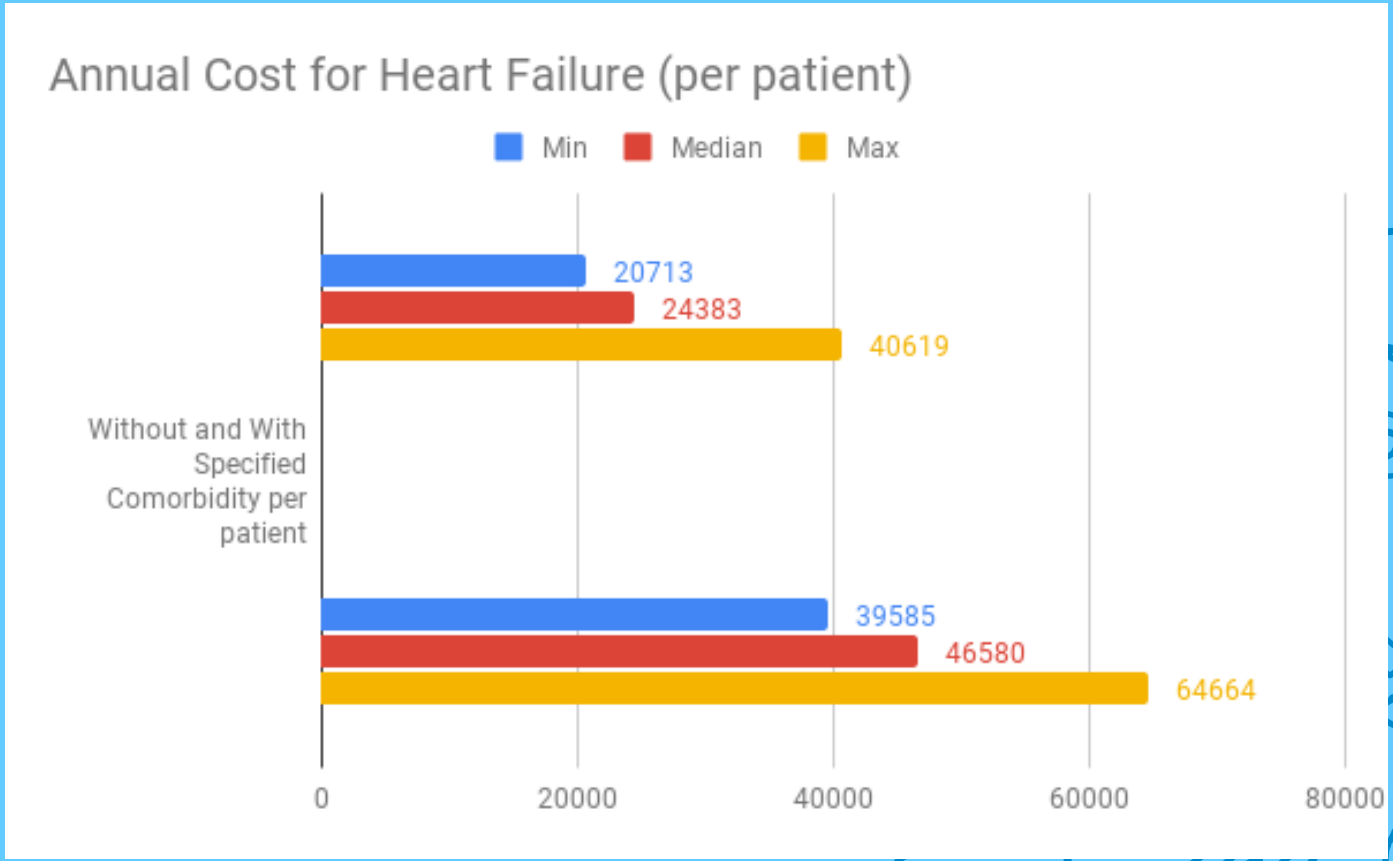
Used to create tables and charts.

Tableau

Create visualization of data set.

Cost

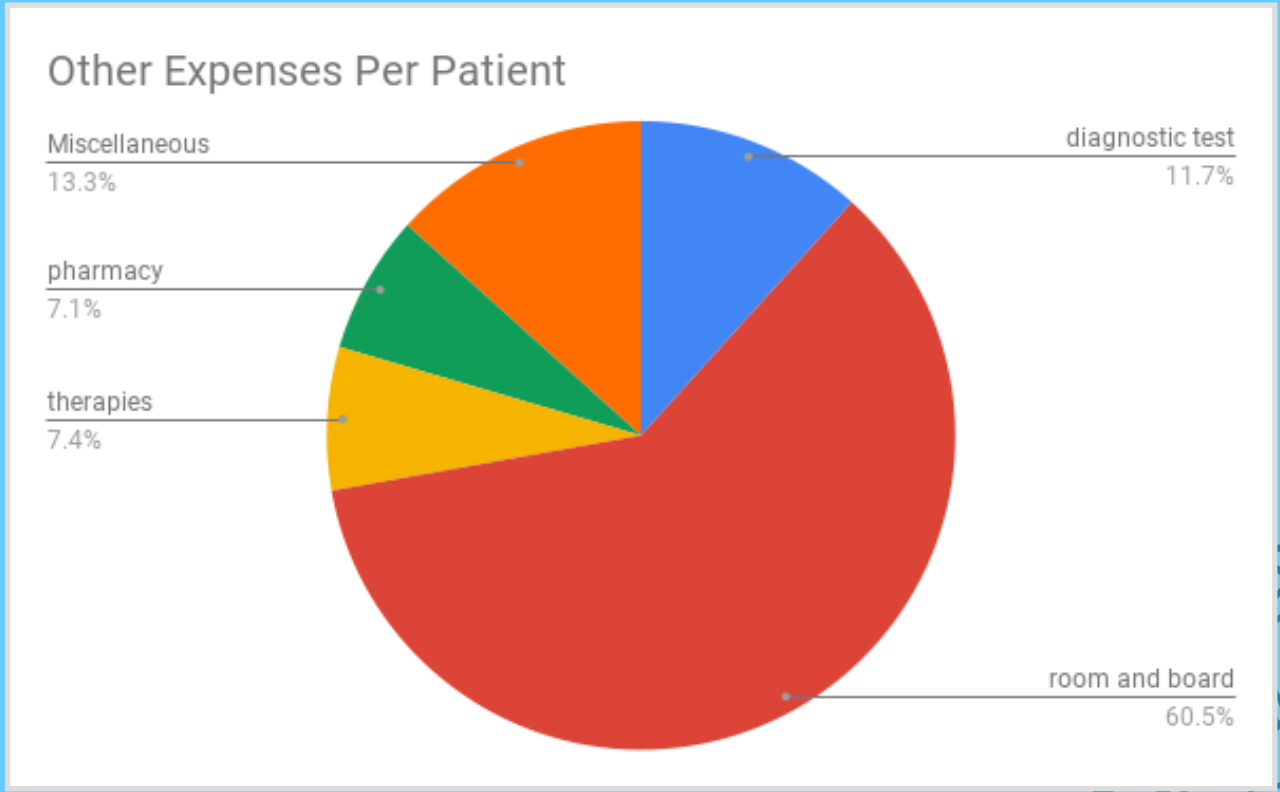
- Price a hospital pays per patient.
- Top: without Comorbidity
- Bottom: with Comorbidity.



Other Expenses

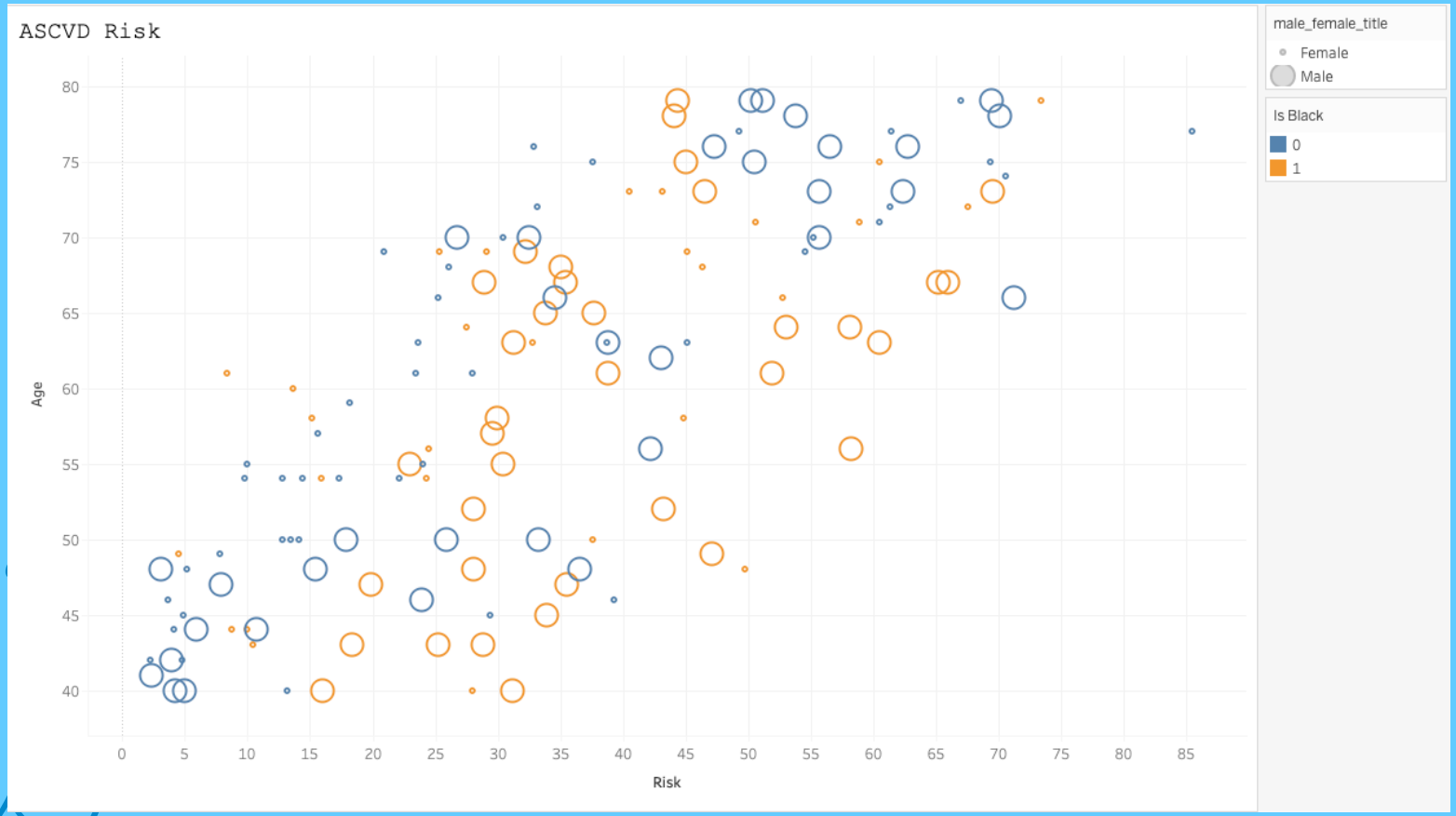
Other expenses a hospital incurs:

- Room and Board
- Diagnostic Test
- Miscellaneous
- Therapies
- Pharmacy




ASCVD Risk
(Atherosclerotic
Cardiovascular Disease)

Results



ASCVD Risk
With filters:

- Smokers
- Diabetic
- Hypertensive

Conclusions

- Recommend building a predictive model will:
    - The doctors can potentially lower the chance for a patient to not incur more medical cost.
    - Lower the amount spent by the hospital to care for patients with heart disease.

Future Work

- Calculate exactly how much money the hospital will save annually.
- Gather more data sets to help build a machine learning model that can help predict heart disease patient.

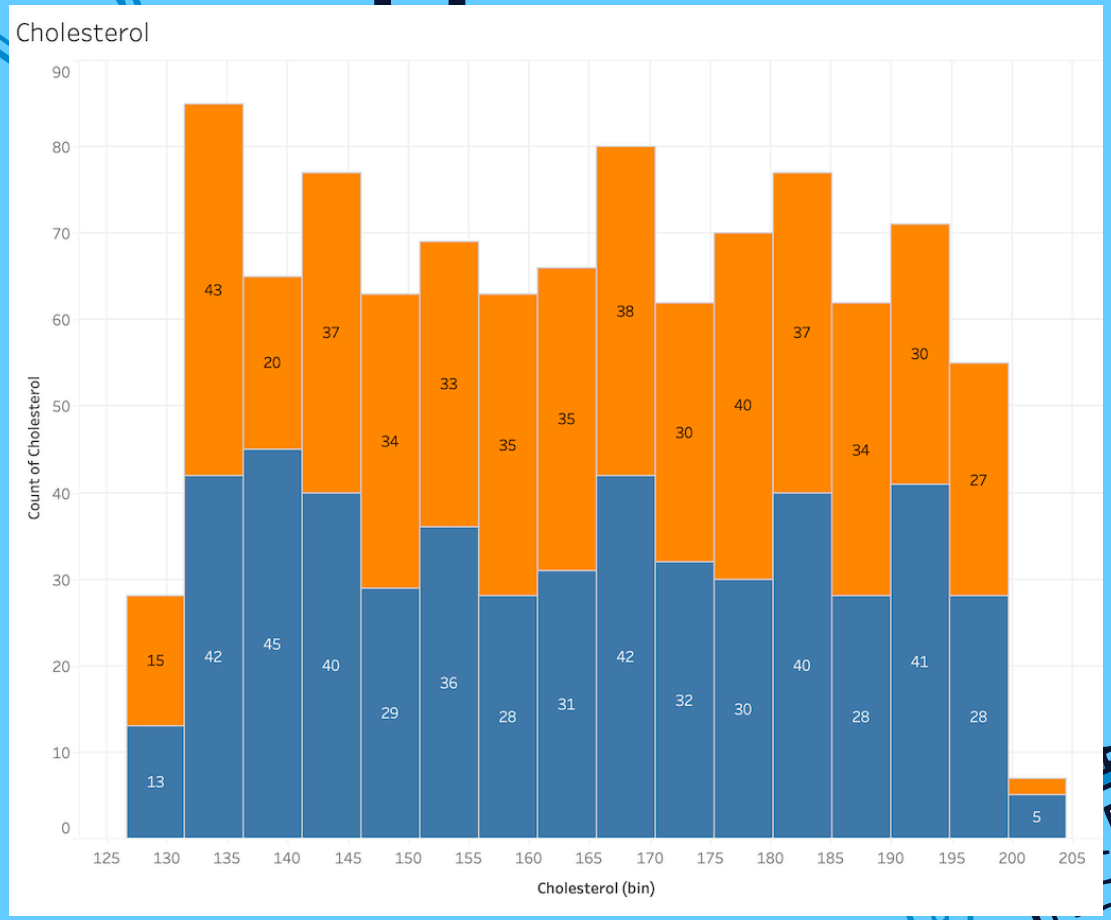
Resources

Websites:

- Kaggle Data set (ASCVD)
- Heart Disease Facts
- A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014–2020)

Thank You

Appendix


Cholesterol
Count of Cholesterol
Cholesterol (bin)
0
10
20
30
40
50
60
70
80
90
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
15
43
20
37
34
33
35
35
38
30
40
37
34
30
27
13
42
45
40
29
36
28
31
42
32
30
40
28
41
28
5

Appendix

ASCVD (%)

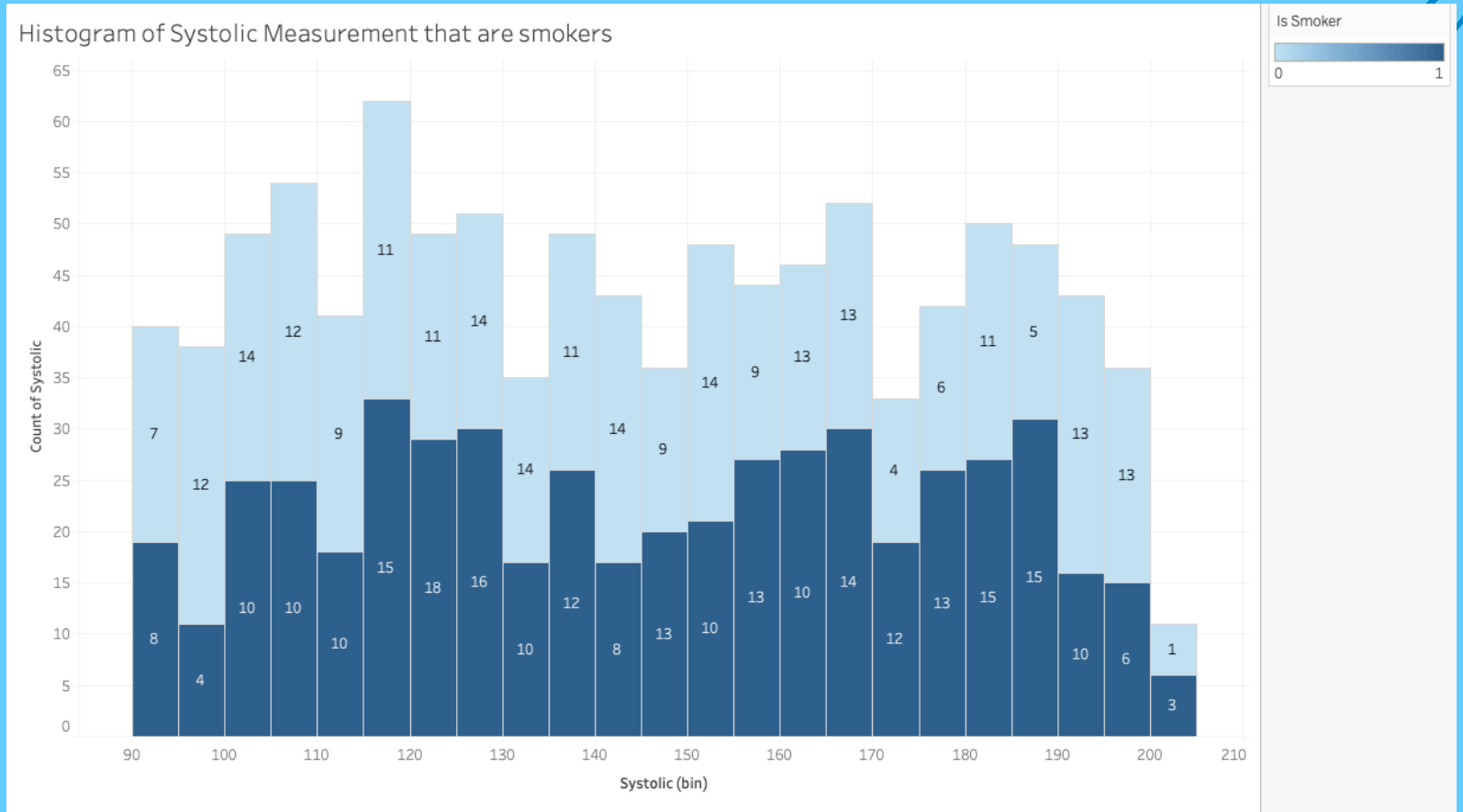
Low - Borderline

0 to 4.9 - Low
5 to 7.4 - Borderline

Intermediate - High

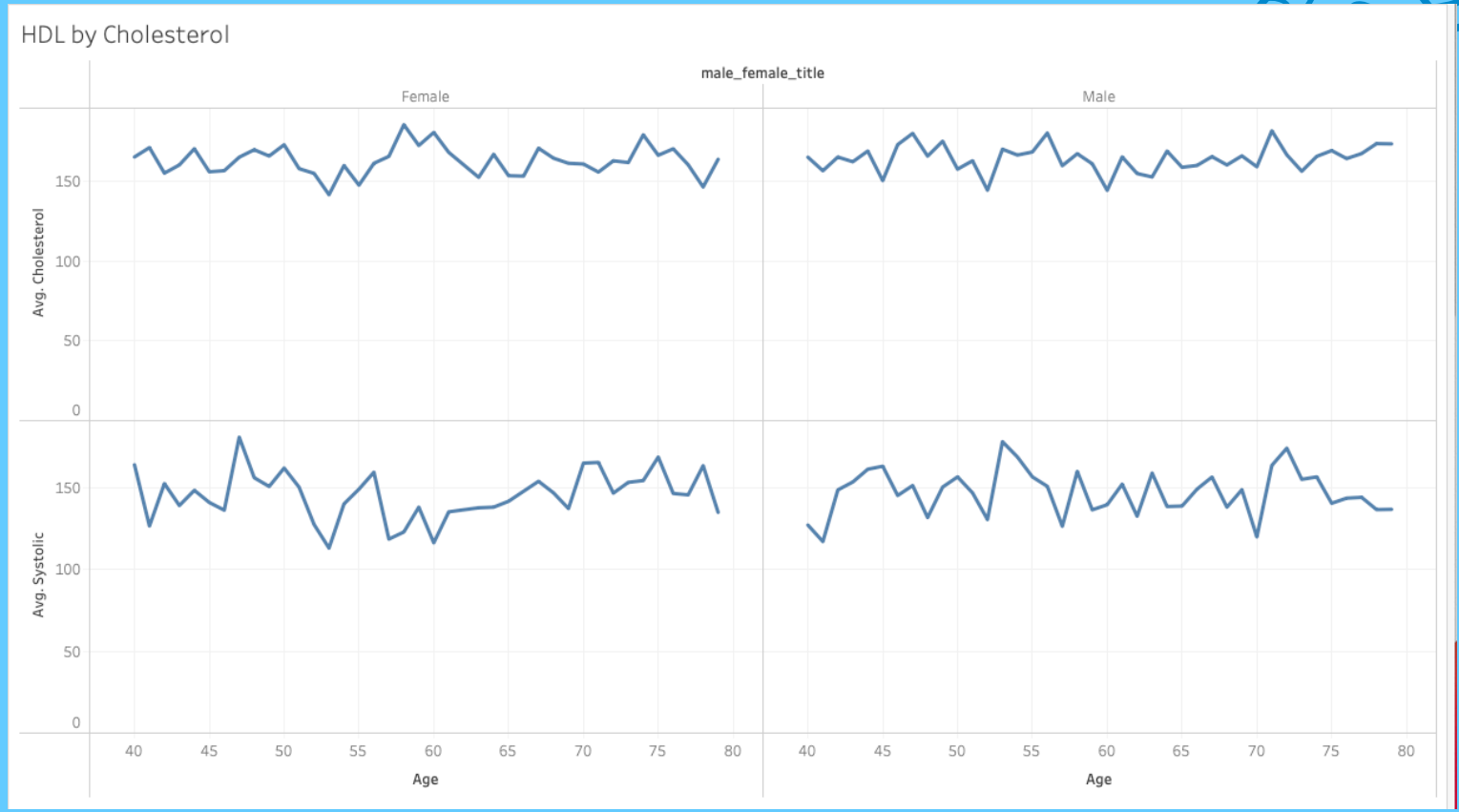
7.5 to 20 - Intermediate
Greater than 20 - High

Appendix

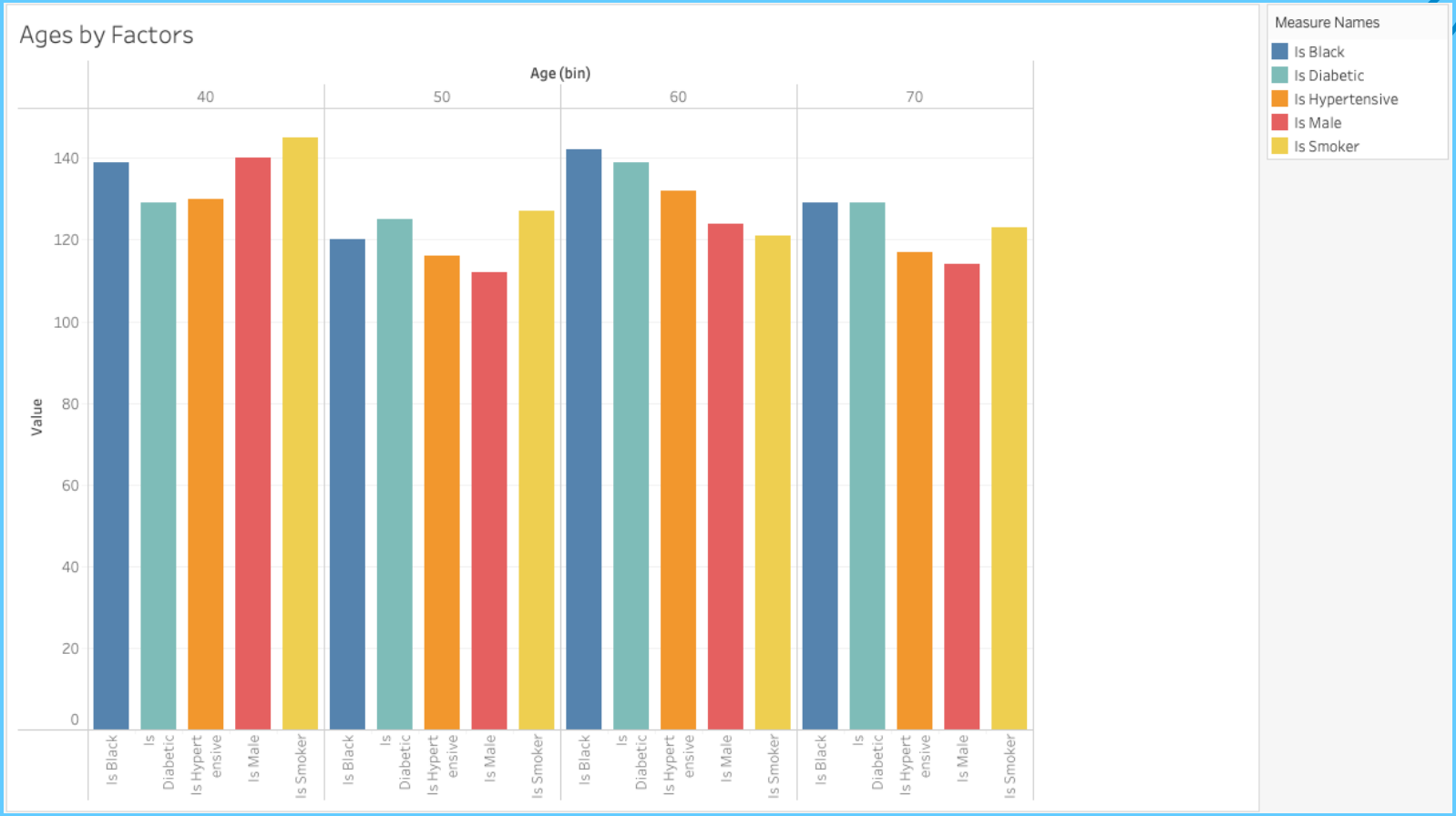

Histogram of Systolic Measurement that are smokers
Count of Systolic
Systolic (bin)
Is Smoker
0
1
65
60
55
50
45
40
35
30
25
20
15
10
5
0
90
100
110
120
130
140
150
160
170
180
190
200
210
7
12
14
12
9
11
11
14
14
11
14
9
14
9
13
13
4
6
11
5
13
13
1
8
4
10
10
10
15
18
16
10
12
8
13
10
13
10
14
12
13
15
15
10
6
3

Appendix

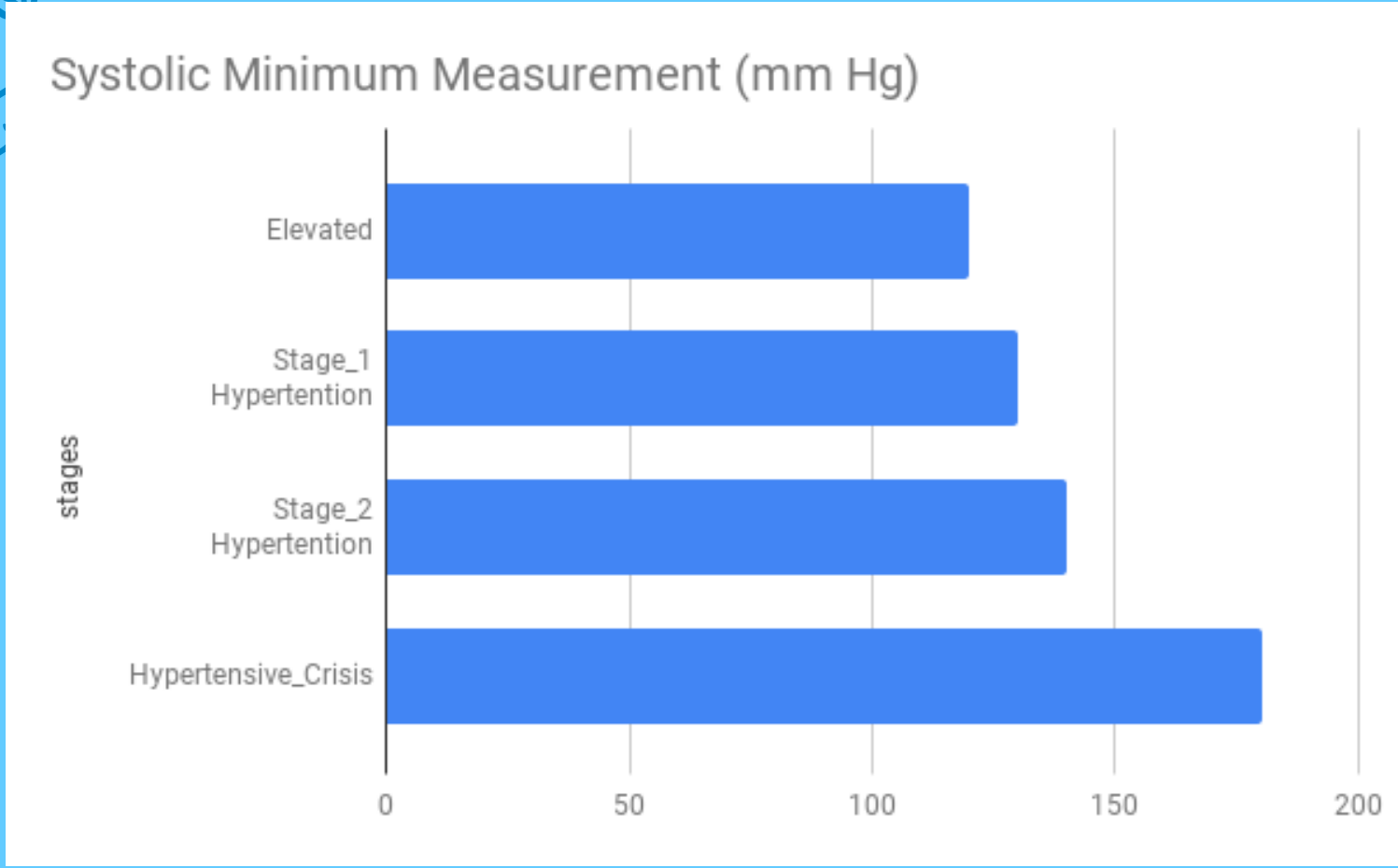
With filter for smokers



Appendix


Ages by Factors
Age (bin)
40
50
60
70
Value
0
20
40
60
80
100
120
140
Is Black
Is Diabetic
Is Hypertensive
Is Male
Is Smoker
Measure Names
Is Black
Is Diabetic
Is Hypertensive
Is Male
Is Smoker

Appendix


Systolic Minimum Measurement (mm Hg)
stages
Elevated
Stage_1 Hypertention
Stage_2 Hypertention
Hypertensive_Crisis
0
50
100
150
200