Data Mining& Exploratory Data Analysis for Payor Data Deeper Dive Review

Zixi Liu, Urban Spatial Analytics, UPenn

Jan 8th, 2020

Exploratory Data Analysis: Payor Data Variables for Reference

1. Data Review

   Payor data has two datasets-readmitted entries and not-readmitted entries. Readmitted dataset has **118883** entries and **79** variables and not-readmitted dataset has **168980** entries and **79** variables.

Adm Date
Admit Pract Name
Admit Pract NPI
APR-DRG Code
APR-DRG Name
APR-DRG SOI
APR-DRG Wt
Charges
Check-in Source
Check-in Type
Days to Readmission
(Index = start of new patient)
Discharge Date
Discharge Status Code
Discharge Status Name
Encounters
Entity Name
Facility of Origin
Fixed Dir Costs
Insured ID Number
Length of Stay
MDC Code
MDC Descr
MSDRG Code
MSDRG Name
MSDRG Wt
Newborn Flag
Pat Home Country Name
Pat Home State Code & Name
Pat Home Zip Code & Name
Patient Account Number
Patient Age
Patient Employer Name
Patient Full Name
Patient MEI Code
Patient Payment
Patient Sex
Patient SSN/National ID
Patient Type
Patient Type Name
PatientDateofBirth
PatientMedRecNo
Posted Denial_Downgrade Only
Posted Denials_Non Downgrade
Pract of Rec Full Name
Pract of Rec NPI
Primary Payor Detailed Act Pmt
Primary Payor Detailed Exp'd Pmt
Primary Payor Financial Class
Primary Payor Grouping
Primary Payor Plan ID
Primary Payor Plan Name
Prin Clin Proc Code
Prin Diagnosis Code
Prin Diagnosis Name
Princ Clin Proc Name
Readmission Pts
Registr Point of OriginName
Sec Diagnosis Code
Sec Diagnosis Name
Secondary Payor Code
Payment Act Pmt
Secondary Payor Plan Code
Secondary Procedure Code
Secondary Procedure Name
Service Line 1 Description
Service Line 2 Description
Service Line 3 Description
Third Diagnosis Code
Third Diagnosis Name
Total Account Balance
Total Cost
Total Fixed Cost
Total Payor Detailed Act Pmt
Total Variable Cost
Transportation
Variable Dir Labor
Variable Dir Supply

Notes:

APR-DRG Code:
Medicaid DRG codes
Check-in Source:
How patients come to Penn
Check-in Type:
ED, Elective, etc.
Discharge Status Name:
Home health service etc.
MSDRG Code:
Medicaid version code
Patient MEI Code:
Link to provider services
Posted Denial_Downgrade Only:
Internal denials (billed as outpatient)
Posted Denials_Non Downgrade:
Payor Denial

Exploratory Data Analysis: Collapsing Payor Data

2. Creating new variables

We define four new variables for analysis:

Read: Readmitted or not, Responses: Y/N

Read30Days: entries <= 30 days Readmission, Responses: Y/N

Read7Days: entries <= 7 days Readmission, Responses: Y/N

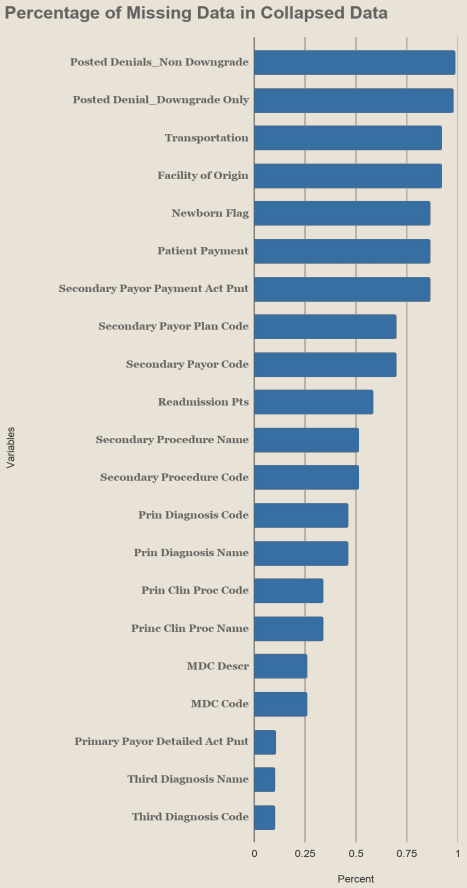
NumOfRead: Number of Readmissions (0 for not readmitted)

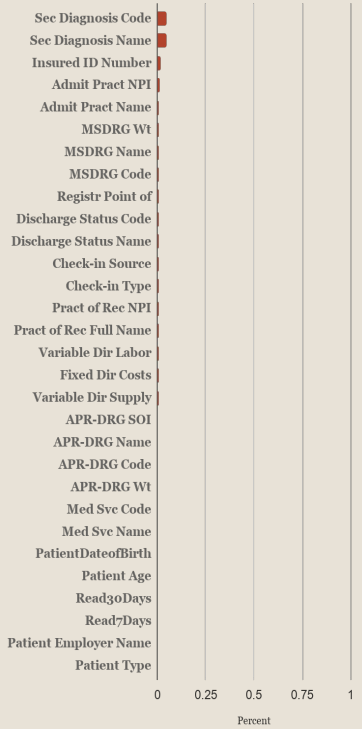
3. Collapsing Datasets

We collapse readmitted and not-readmitted datasets and collapsed data has 287863 entries and 83 variables. Only Readmitted Data has Primary Payor Grouping.

We convert data types and clean the data, remove the $ etc. We found variable Days to Readmission contains negative values and requires revision.

Variables colored in blue have > 10% missing data. Variables colored in red have <= 10% missing data.



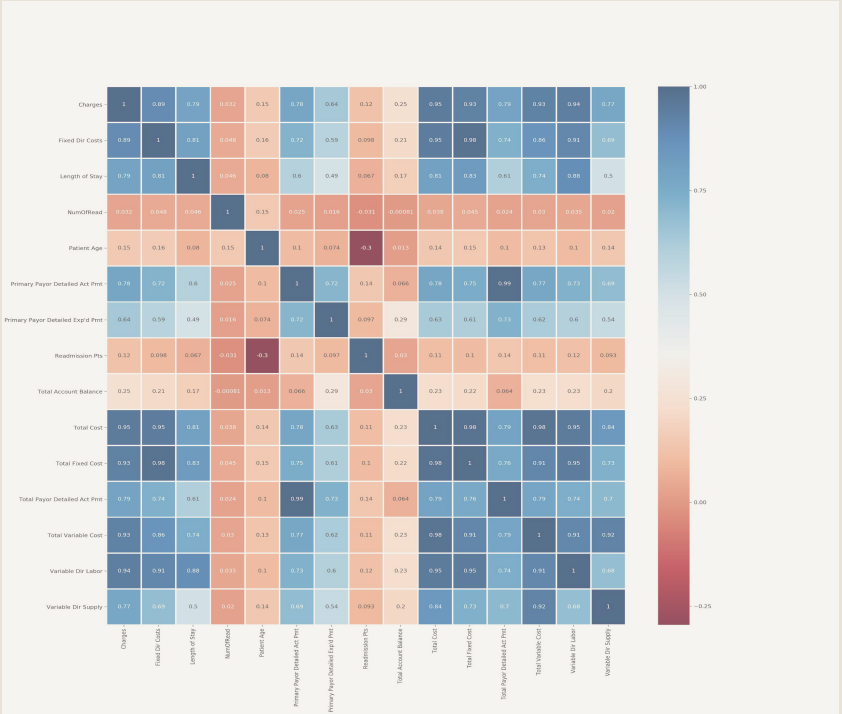


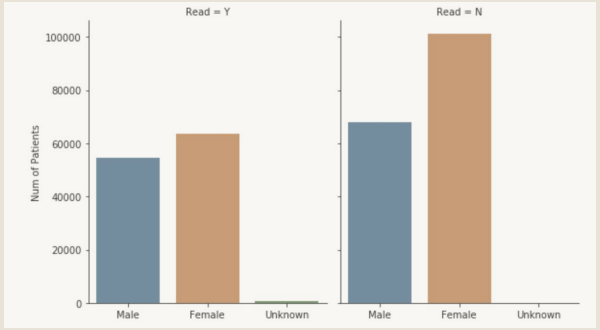
Exploratory Data Analysis: Collapsing Payor Data& Correlation Map

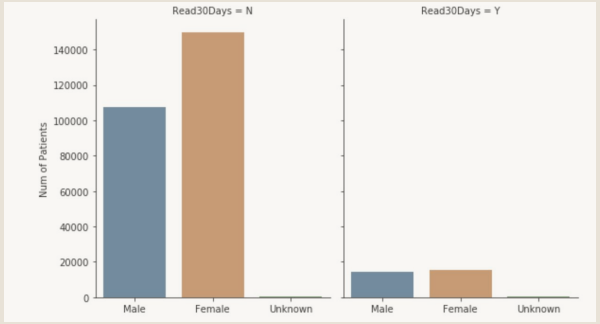
4. Correlation Map

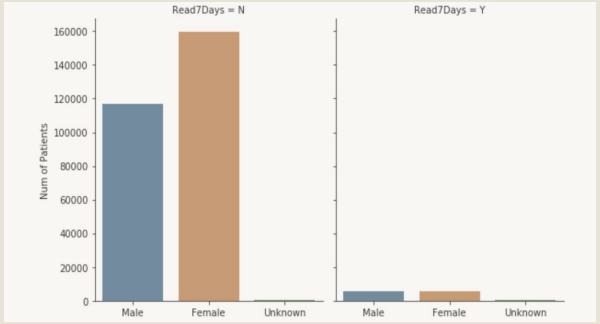
From the correlation map, Patient Age has a negative relationship to Readmission Pts; costs and payments have strong relationships with length of stay.

- In collapsed data, 118883 entries readmitted, 168980 entries not readmitted.
- Of all entries readmitted, 30292 entries readmitted within 30 days, 88495 not readmitted within 30 days, 96 have negative values in days to readmission. (Need further examination of negative values)
- Of all entries readmitted, 11228 entries readmitted within 7 days, 107559 entries Not readmitted within 7 days, 96 entries have negative values.



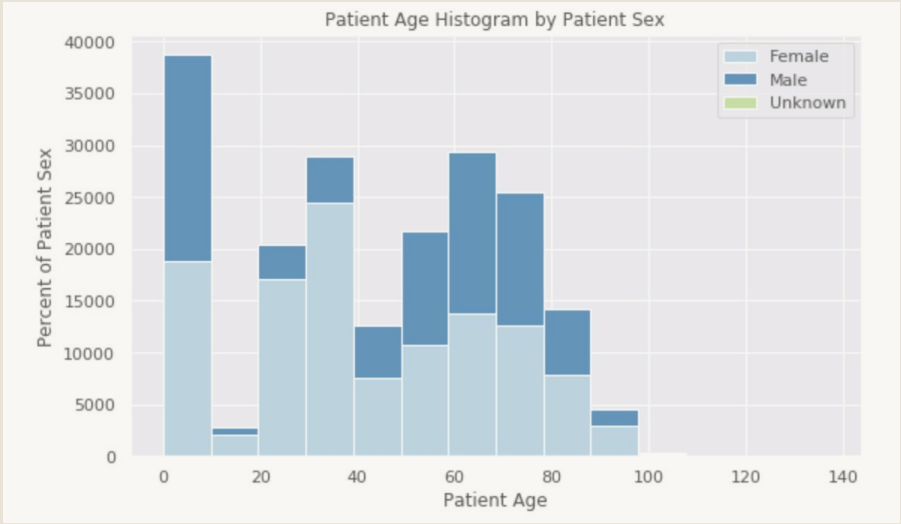




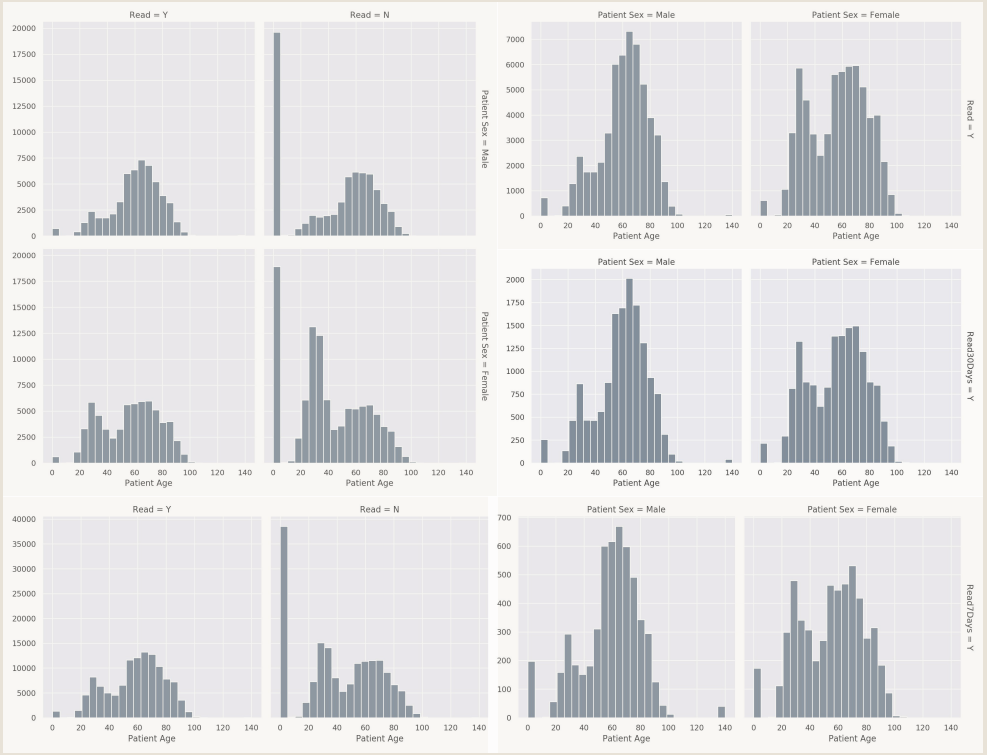


Exploratory Data Analysis: Patient Age& Sex

- 164830 entries are female, 122271 entries are male, 762 entries are unknown.
- More female entries than male in terms of readmission, readmission within 30 days, and readmission within 7 days.
- After dropping duplicated readmission entries and leaving only patients, there are 199055 patients in total. Approximately same counts of female and male aged 0-10 and 50-80; whereas more female than male aged 10-50 and 80-100.



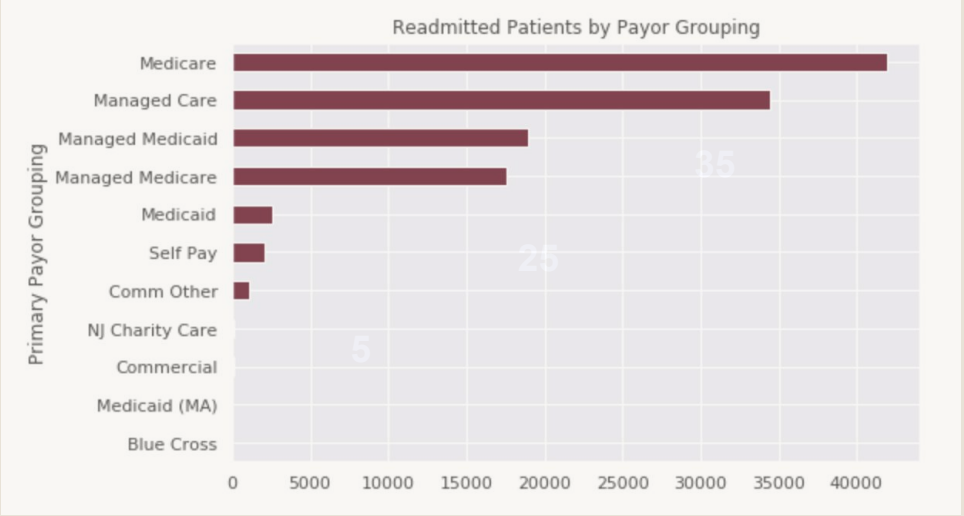
Exploratory Data Analysis: Patient Age& Sex

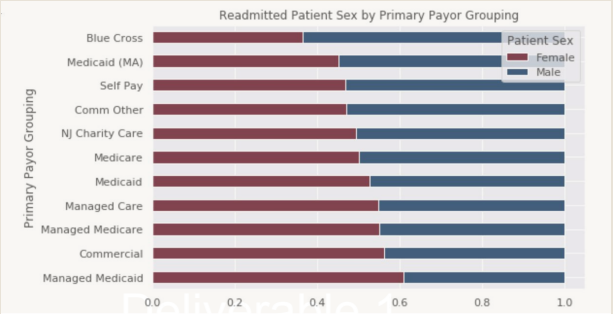


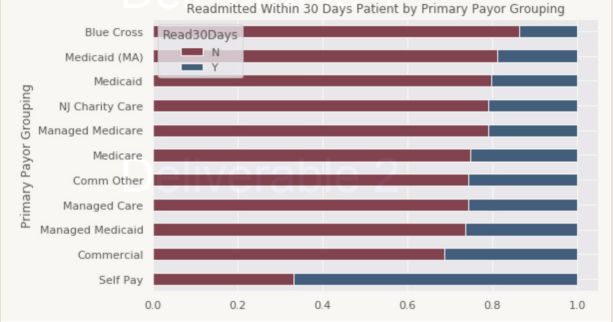
- Of all readmitted data entries, male entries are more centered around the age of 50-80 than female entries.
- From the histogram, there are more counts of readmitted patients aged 50-80.

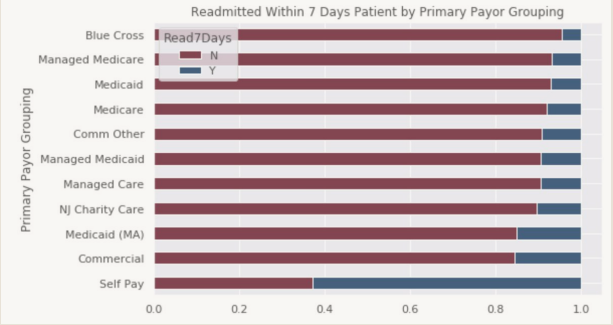
Exploratory Data Analysis: Patient Age& Sex& Primary Payor Grouping

- Of the histogram of patients readmitted within 30 days, more counts of readmitted patients aged 50-80.
- Of the histogram of patients readmitted within 7 days, more counts of readmitted patients aged 50-80.



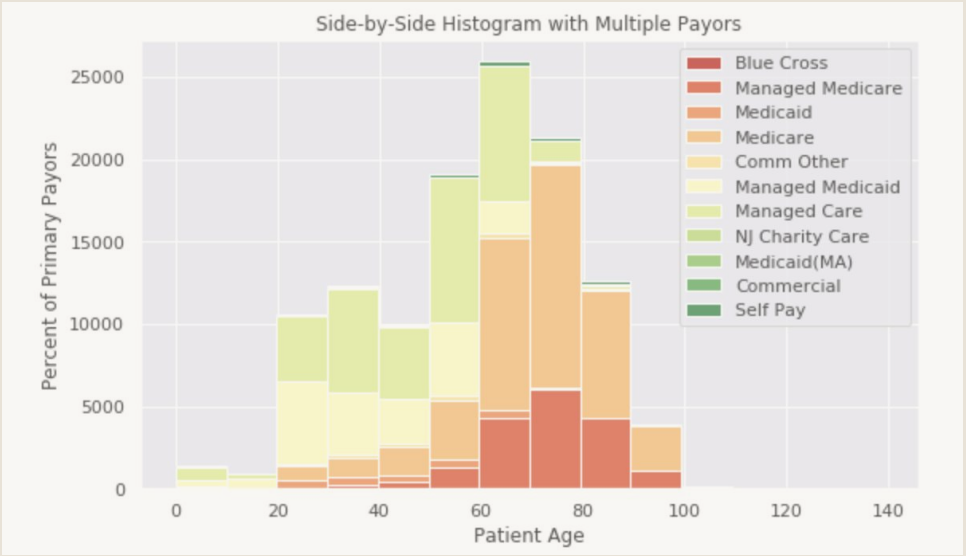






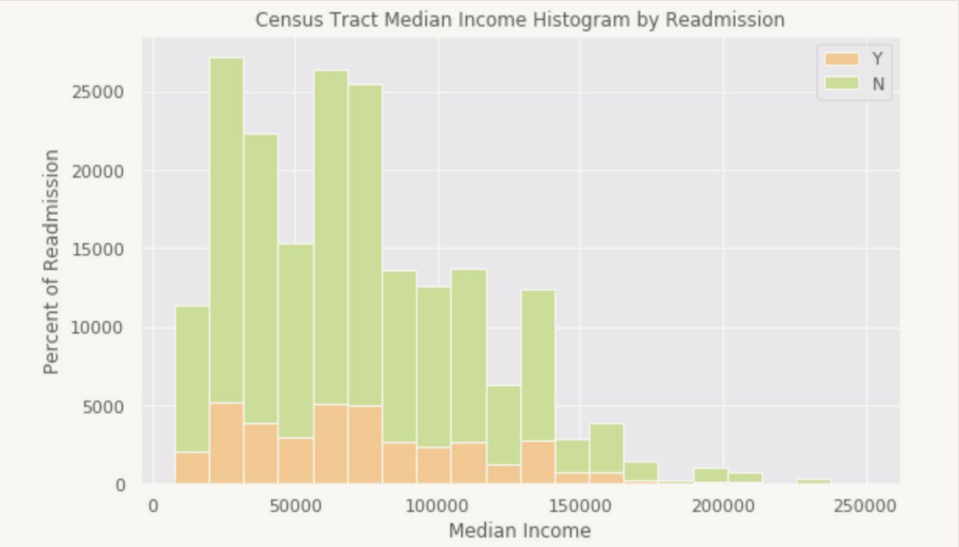
Exploratory Data Analysis: Primary Payors Grouping

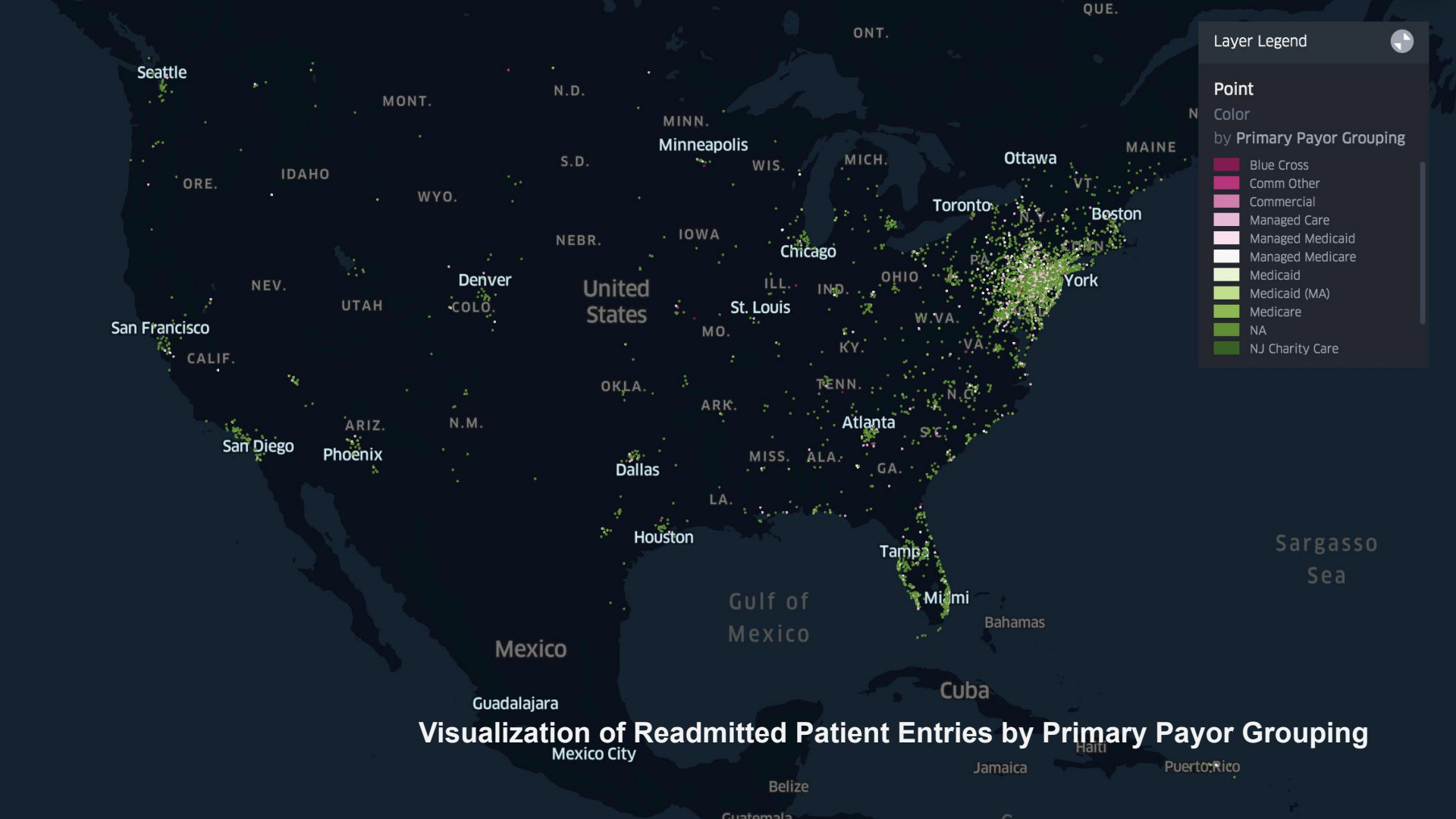
- More entries using Medicare, Managed Care, Managed Medicaid, Managed Medicare, Medicaid.
- Of entries readmitted within 30 days, Self-payed and Commercial has the highest ratio of readmission; Blue Cross, Medicaid(MA) and Medicaid has lowest ratio of readmission.
- Of entries readmitted within 7 days, Self-payed and Commercial has the highest ratio of readmission; Blue Cross, Managed Medicare and Medicaid has lowest ratio of readmission.
- Of all readmitted entries, Medicare has more entries aged 60-90.
- Managed Care has more entries aged 50-70.
- Managed Medicaid has more entries aged 20-60.
- Managed Medicare has more entries aged 60-90.



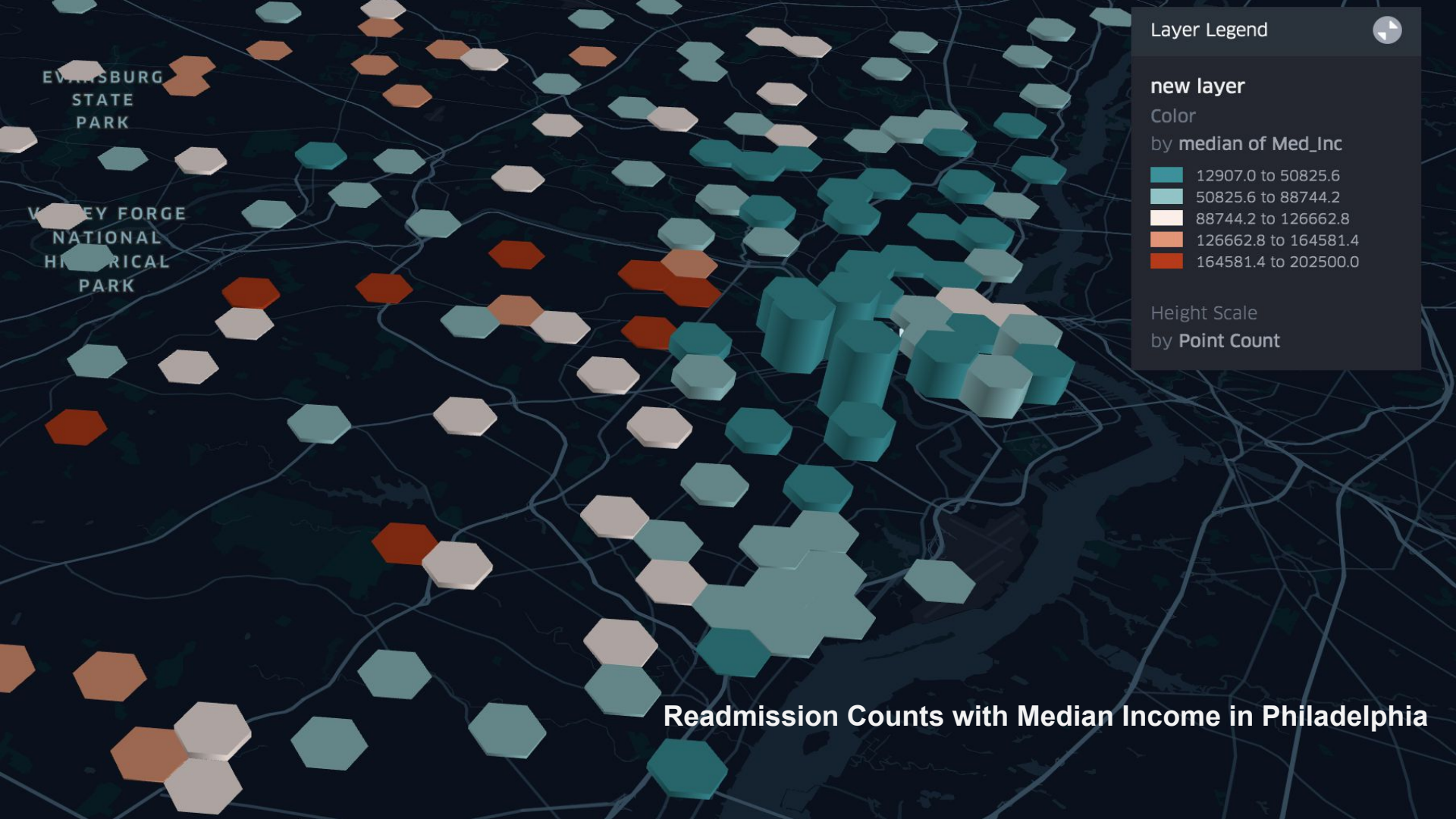
Exploratory Data Analysis: Spatial Analysis

- Parse Patient Home Zip Code& Name by Zip code and Name
- Join Longitude and Latitude data by Zip Code
  Zipcode Websites for Reference:
  https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/
  https://gist.github.com/erichurst/7882666
- Merge joined data with Census Tract Data from 2018.
  Create five new variables: Total Population, Median Income, White Population, Median Age, Origin Tract, Percentage of White Population.
- 1217 entries has missing data for Patient Home Zip Code & Name, listed as ITEM NOT FOUND or Air Force.
- From visualization of median income vs. Readmission, median income lower than $80,000 has more readmission.
- Visualize collapsed data by Primary Payor Grouping.





Visualization of Readmitted Patient Entries by Primary Payor Grouping


EVANSBURG
STATE
PARK
NATIONAL
PARK
Layer Legend
new layer
Color
by median of Med_Inc
12907.0 to 50825.6
50825.6 to 88744.2
88744.2 to 126662.8
126662.8 to 164581.4
164581.4 to 202500.0
Height Scale
by Point Count
Readmission Counts with Median Income in Philadelphia

Data Files for Reference

Collapsed.csv: Collapsing Redmitted and Not Readmitted Data, Creating four new variables.

MergedCodes.csv: Merging Collapsed Data with Lat and Lon of Zip Codes.

JoinPA.csv: Focus on data in Philadelphia.