Saransh Rakshak

Digital Humanities 100
Summer 2021
Prof. Adam Anderson

Analysis of Opiate Prescription Trends in the United States

Project Description

Opiates are potent drugs that are widely used in medicine for the treatment of pain. Unfortunately, these drugs are also highly efficient at activating the ‘pleasure center’ of the brain.

- Due to this property, opiates have a high potential for abuse and addiction when prescribed, and are aptly dispensed as controlled substances.
    - Higher potency opiate prescriptions such as Oxycodone and Fentanyl are schedule II (highly monitored)
    - Lower potency opiates such as Tramadol and Codeine are less strictly monitored and controlled, being schedule IV and III.

- This project will attempt to classify areas of the United States by both average income of a region as well as type of region, and the amount of opiates prescribed by doctors in that area.

Questions

- Primary Question: Are opioids prescribed more frequently in certain states or regions, dependent on income?
  - B: Are prescriptions less common in states that place additional limitations on doctors for prescribing opioids such as California?
  - C: Is Long-Acting opioid treatment therapy used more often in higher income areas? (note: patients with access to long-acting medications may have a higher potential for abuse and addiction)
  - D: How have rates of prescriptions in the US changed over the years?

Data Selection (CMS)

- Data regarding opioid prescriptions is provided by the **Center for Medicare and Medicaid Services**, which can be accessed here:
  - https://data.cms.gov/Medicare-Part-D/Medicare-Part-D-Opioid-Prescriber-Summary-File-201/j2ra-95gh
- Original Data Information
  - Provided By: CMS
  - Contact Email: MedicareProviderData@cms.hhs.gov
  - Bureau Code: 009:38
  - Program Code: 009.000
  - Last Updated: February 8, 2021
  - Number of Rows: 1,204,935
  - Number of Columns: 11

CMS Dataframe

	NPI	last_name	first_name	zip_code	state	doc_spec	tot_presc_cnt	op_cnt	op_rate	LA_op_cnt	LA_op_rate
0	1003000126	ENKESHAFI	ARDALAN	21502.0	MD	Internal Medicine	492	13.0	0.03	NaN	NaN
1	1003000142	KHALIL	RASHID	43623.0	OH	Anesthesiology	1818	891.0	0.49	143.0	0.16
2	1003000167	ESCOBAR	JULIO	89403.0	NV	Dentist	77	NaN	NaN	0.0	NaN
3	1003000282	BLAKEMORE	ROSIE	37243.0	TN	Nurse Practitioner	100	0.0	0.00	0.0	NaN
4	1003000407	GIRARDI	DAVID	15825.0	PA	Family Practice	2766	22.0	0.01	NaN	NaN

- **NPI**: National Provider Identifier for Doctor Identification.
- **last_name**: Prescribing Doctor's Last Name.
- **first_name**: Prescribing Doctor's First Name.
- **zip_code**: **Zip code of the doctor's location**
- **state**: **Prescribing doctor's location.**
- **doc_spec**: Specialty of Doctor.
- **tot_presc_cnt**: Total amount of prescriptions by the doctor (inclusive of non-opioids).
- **op_cnt: Amount of opioid prescriptions by the doctor.**
- **op_rate: Changes in the amount of opioid prescriptions over time by each individual doctor.**
- **LA_op_cnt: Amount of Long-Acting opioid prescriptions by the doctor.**
- **LA_op_rate: Changes in the amount of Long-Acting opioid prescriptions over time by each individual doctor.**

Data Selection (IRS)

- Data regarding average incomes in different area codes is originally provided by the **Internal Revenue Service** (IRS) , which can be accessed here:
    - https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2016-zip-code-data-soi
- Data that I will be using has been cleaned and posted for public access by Jon Loyens on data.world, which can be accessed here:
    - https://data.world/jonloyens/irs-income-by-zip-code
- Original Data Information
    - Provided By: Jon Loyens
    - Contact Address: @jonloyens (twitter)
    - Last Updated: June 16, 2016
    - Number of Rows: 27,790
    - Number of Columns: 11

IRS Dataframe

	state	zip_code	num_tax_ret	zip_agi	zip_avg_agi	ret_w_total	sum_zip_income	zip_avg_income	num_ret_taxable	taxable_amt	avg_taxable
0	AL	0	2022380	105089761	51.963410	2022380	106420533	52.621433	1468370	67850874	46.208295
1	AL	35004	4930	255534	51.832454	4930	258024	52.337525	4020	163859	40.760945
2	AL	35005	3300	128387	38.905152	3300	129390	39.209091	2440	70760	29.000000
3	AL	35006	1230	58302	47.400000	1230	58585	47.630081	940	36341	38.660638
4	AL	35007	11990	643708	53.687073	11990	651350	54.324437	9280	414878	44.706681

- **state**: The state the zip code is located in.
- **zip_code**: Zip code.
- **num_tax_ret**: The number of tax returns filed in 2016.
- **zip_agi**: Adjusted gross income for residents in a zip.
- **zip_avg_agi**: Average gross income for residents in a zip.
- **ret_w_total**: Number of income tax returns with total income.
- **sum_zip_income**: Summation of all incomes of residents in a zip code.
- **avg_income**: Average income of residents in a zip code, in the thousands. (i.e. 52.3 = $52,300 amount earned, on average, by residents in a shared zip code)
- **num_ret_taxable**: Number of income tax returns with taxable income.
- **taxable_amt**: Summation of the total amount a zip can be taxed.
- **avg_taxable**: Average total taxable income for each area code, in the thousands. (i.e. 40.76 = $40,760 income tax paid by that zip code in 2016)

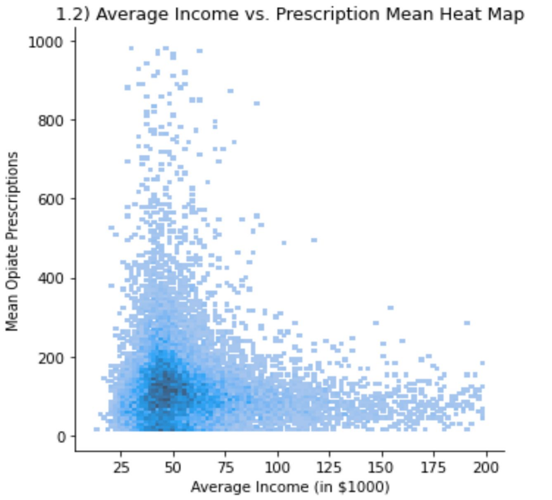
Cleaning and Merging

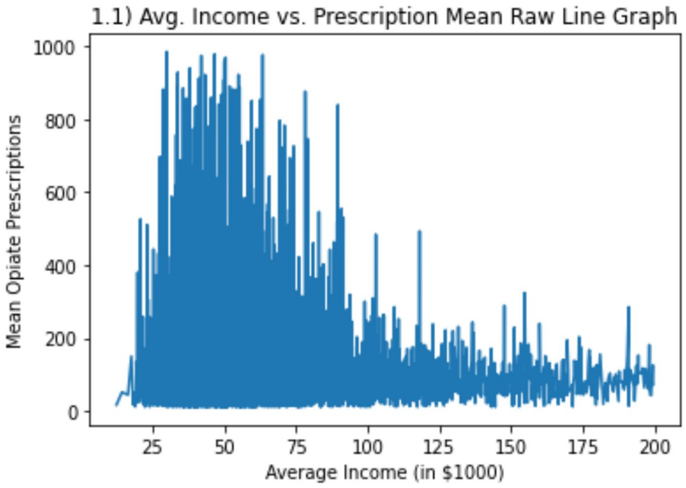
- Dataframe for prescription data and income data was cleaned by removing unnecessary rows/columns and editing corrupted or non-existent values
- Next, the two dataframes were merged, **op_data**, which contains prescription information as well as ZIP, and **income_data**, which contains income information as well as ZIP. We will combine on the ZIP datapoint, and data from income will be merged onto the **op_data** dataframe (**op_data** will remain stable) and mean is taken for activity of each values in zip codes.
- Finalized DataFrame **‘zip_df’**(used for creating visualizations):

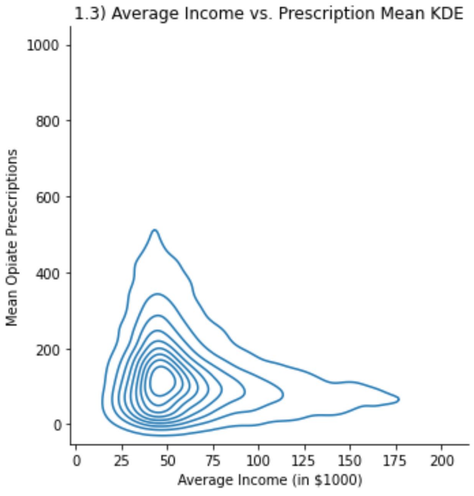
	op_cnt	op_rate	LA_op_cnt	LA_op_rate	zip_avg_income	income_bin
zip_code						
10001	42.121951	0.079268	4.268293	0.043171	155.101676	155
10002	46.020000	0.024200	2.840000	0.043000	46.846786	46
10006	78.333333	0.136667	9.500000	0.028333	181.589407	181
10009	44.833333	0.065833	3.083333	0.014167	72.640410	72
10025	65.891089	0.135248	4.445545	0.069406	133.870043	133

Exploratory Data Analysis 1

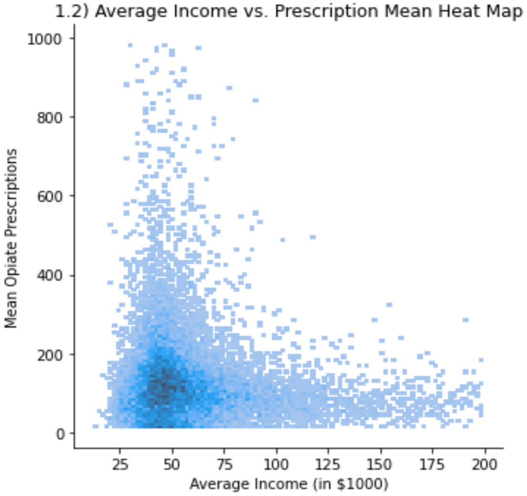
Are opioids prescribed more frequently in certain states or zip codes with higher average income?

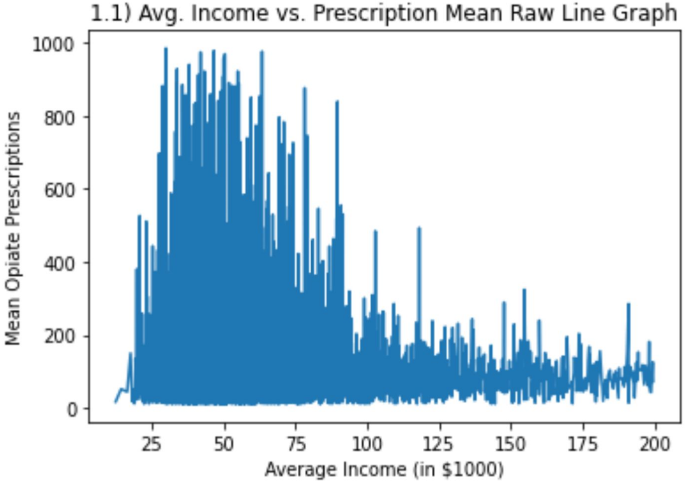


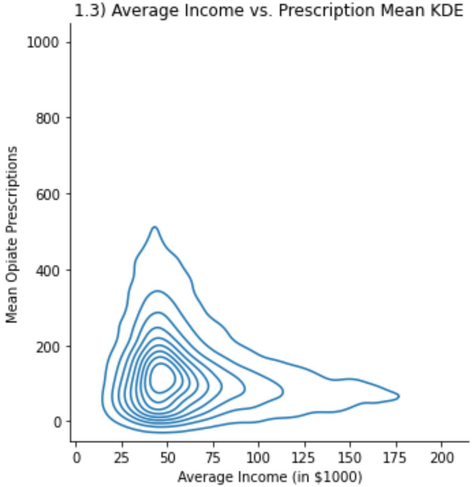




Exploratory Data Analysis 1





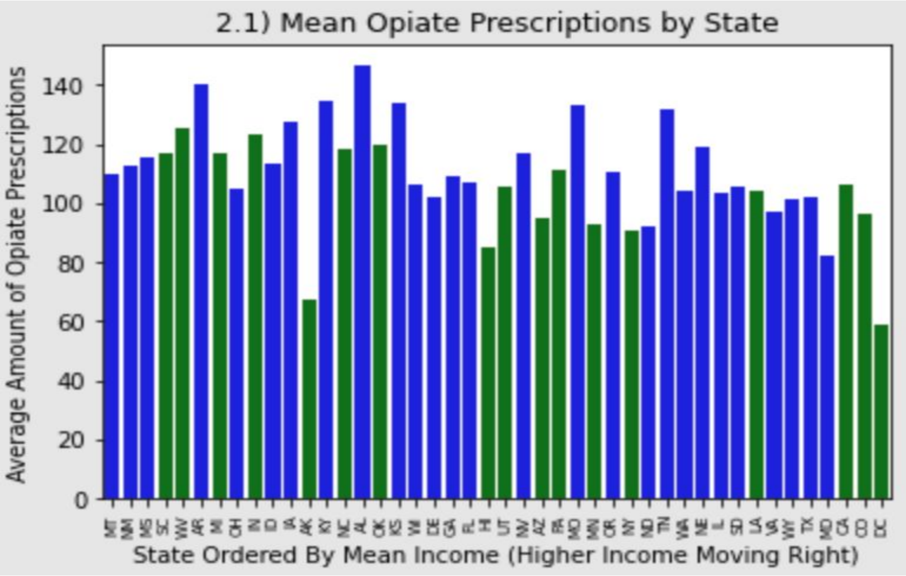


Opioids appear to be less frequently prescribed in regions with a higher average income, as shown by the stretching in fig.1.3 as well as the heat map (fig.1.2). The mean prescription density spreads horizontally and drops steeply and significantly after region incomes of USD 50,000. Thus, we can see as income increases, opioid prescription frequency decreases. There is also a decrease in opioid prescribing for income areas below USD 25,000. One possible explanation for this could be an inability to afford proper healthcare and be able to see a doctor to prescribed opioids- even if they may really need them.

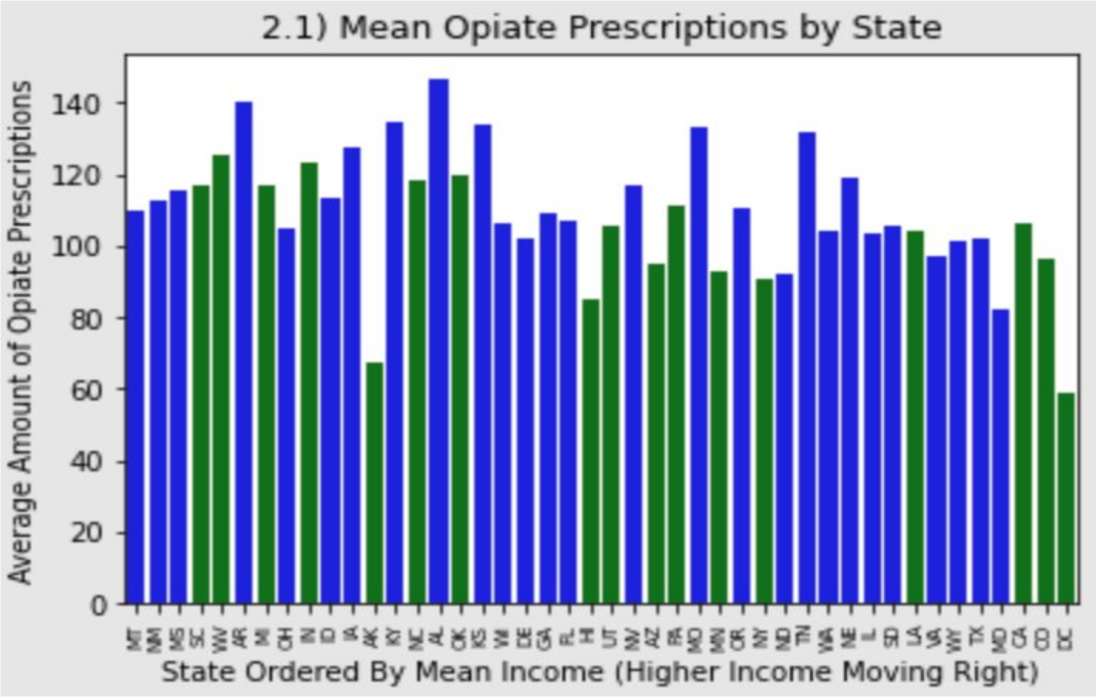
Exploratory Data Analysis 2

Now we will group our data values by state, using the income of all zips in a state to create a column for each state's average income, labeled 'state_avg_income'. Certain states have **additional limitations** regarding prescription. For instance, **California** requires that opioid naive (first time opioid users) must be limited to a 7 day (or less) prescription.

States using additional limitations regarding opioids also include: **Alaska, Hawaii, Colorado, Utah, Oklahoma, Louisiana, Missouri, Indiana, West Virginia, South Carolina, Pennsylvania, New York, Maine, Arizona, North Carolina,** and **New Jersey**. These states are marked in **GREEN**.



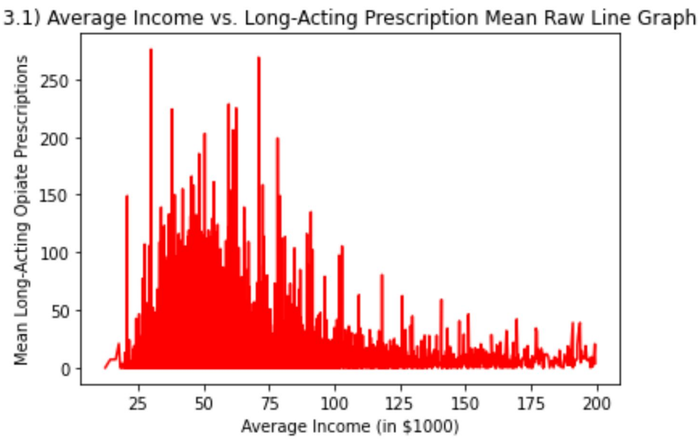
Exploratory Data Analysis 2

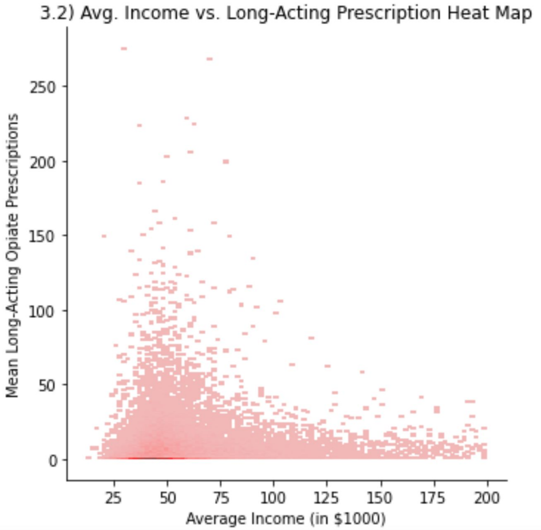


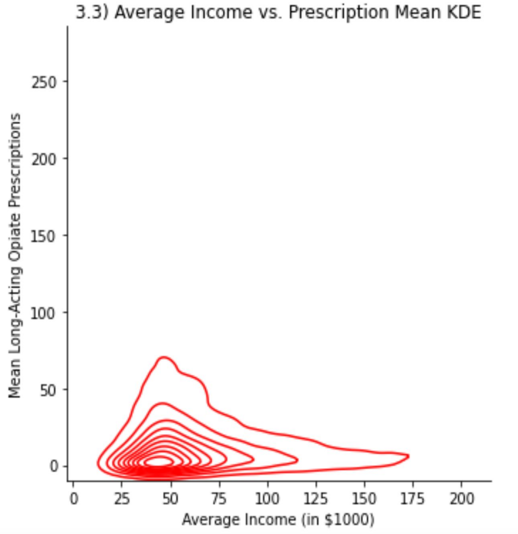
From looking at graph 2.1, cannot definitively conclude that states with higher average incomes have lower frequency of opioid prescriptions. However, it can be observed that states placing additional, more stringent requirements on prescriptions have distinctive decreases in mean opioid prescriptions.

Exploratory Data Analysis 3

Additionally, I want to also find out if **Long-Acting** opioid treatment therapy is used more often in higher income areas.

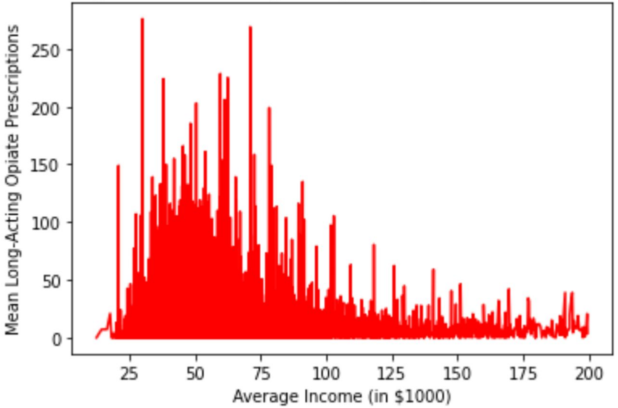




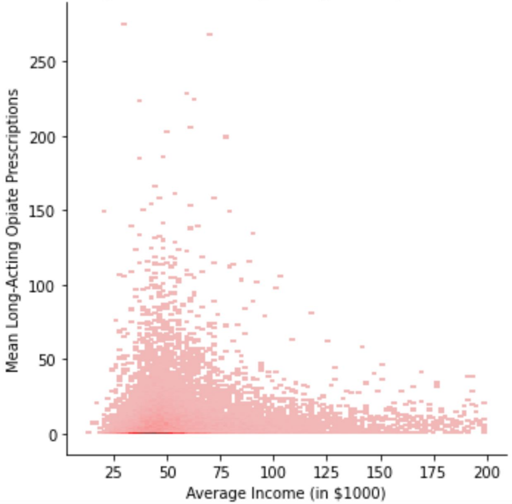


Exploratory Data Analysis 3

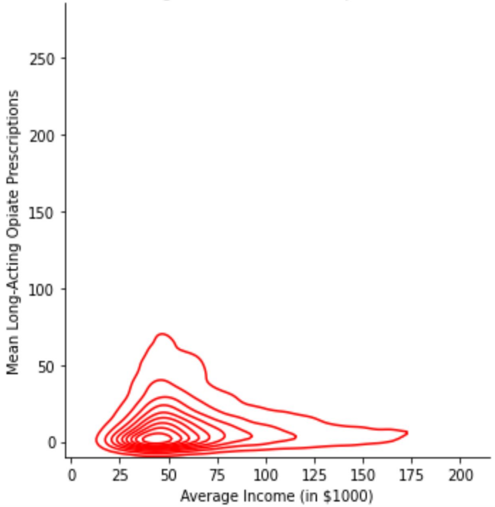
3.1) Average Income vs. Long-Acting Prescription Mean Raw Line Graph



3.2) Avg. Income vs. Long-Acting Prescription Heat Map



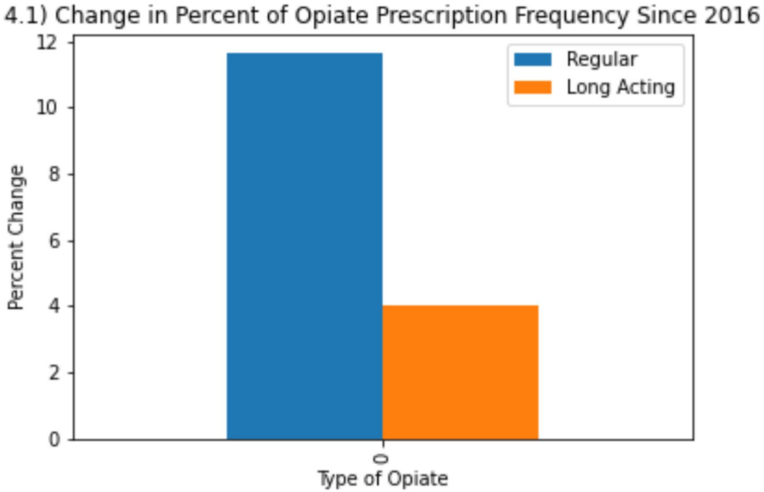
3.3) Average Income vs. Prescription Mean KDE



Although there is similar decrease in frequency as income increases between regular and Long-Acting opioids, it is notable that there is much quicker decreases for short acting. Thus, to some extent, there can be a greater frequency of prescribing LA opioids instead for higher income areas, even though the average amount of opioids prescribed still remains lower for higher income areas. A possible explanation for this is that those in higher income areas may live busier life, and thus doctors may prescribe Long-Acting medication to prevent skipped or forgotten doses.

Exploratory Data Analysis 4

Finally, I would like to see the change in prescriber rates for both regular and Long-Acting opioid prescriptions over the last 5 years. Our DataSet has 2 columns that hold this information, op_rate and LA_op_rate, which show percent changes over the course of the last 5 years for all doctors.



There has been a 11.91% increase in regular acting prescriptions and 4.24% increase in Long-Acting prescriptions for opioids over the last 5 years.

Additional Information

- All information for this project, including python files used to create visualizations, can be found at the following sources below:
  - Github: https://github.com/saranshrakshak/DigHum100
  - GDrive: https://drive.google.com/drive/u/1/folders/1_7ydvkErCFqNI3Dgthyjl_yGayYw2fVR