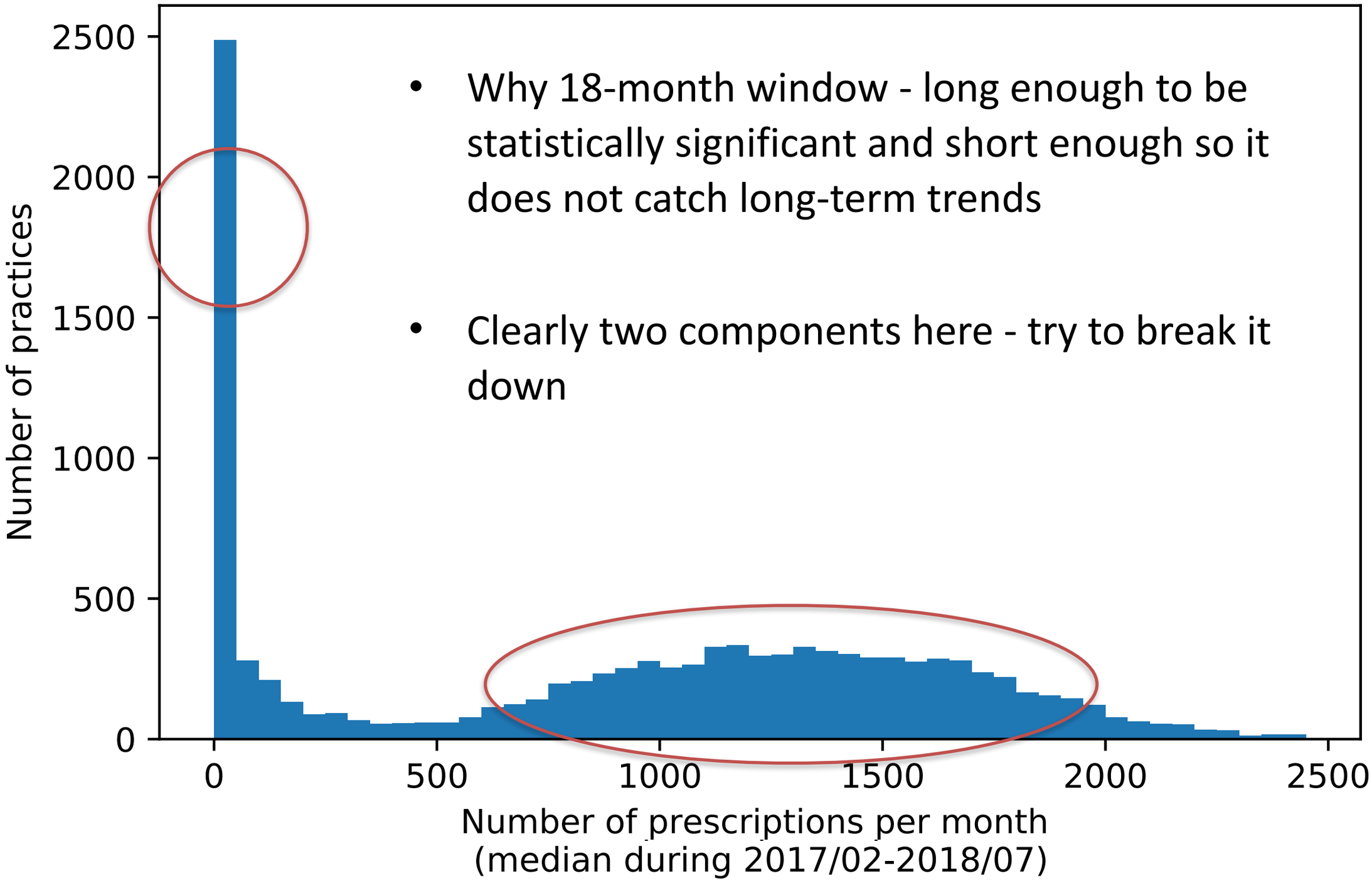
A Statistical Study on GP practice prescribing data in England

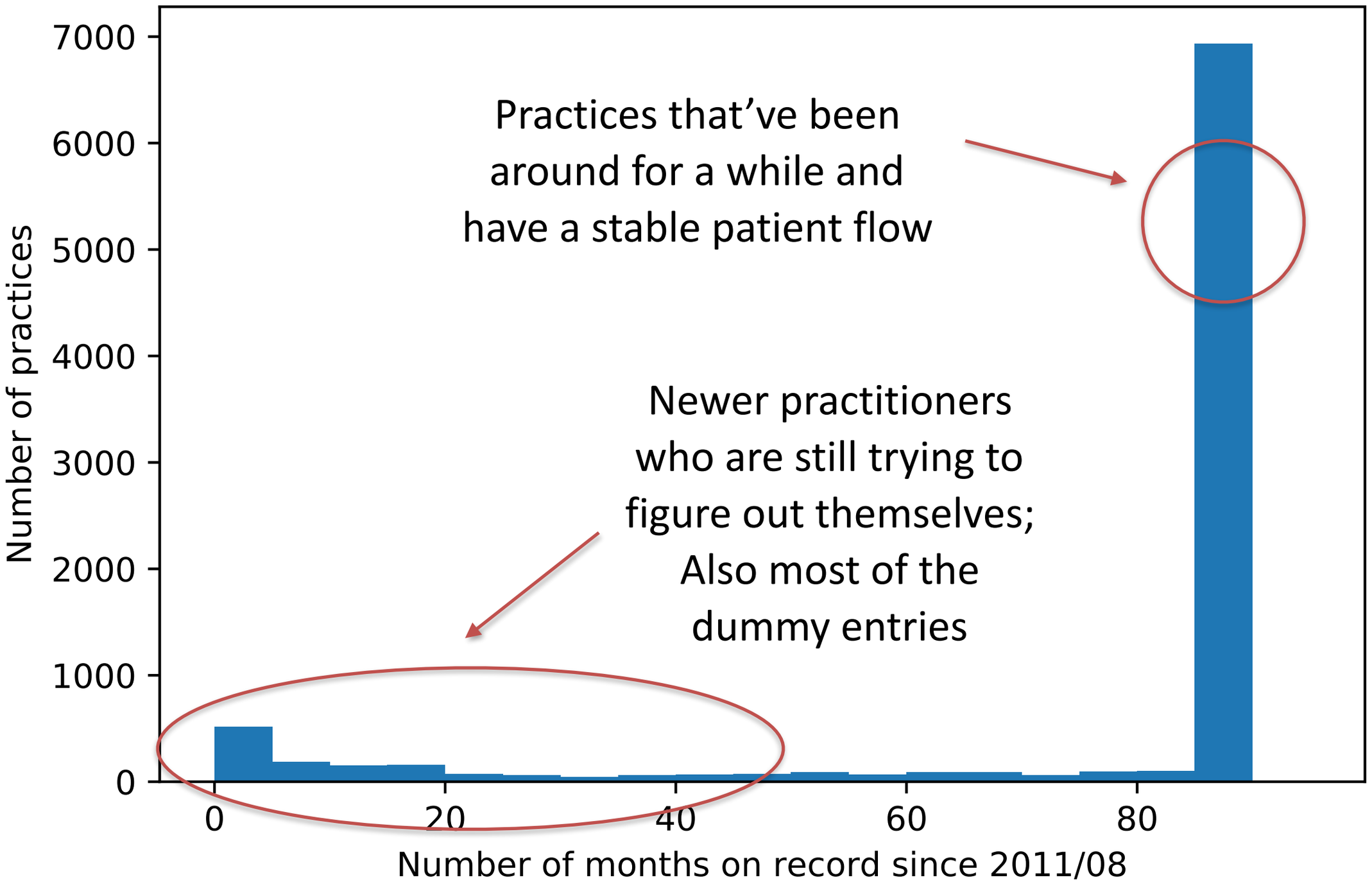
Chen-Chen Zhou

Histogram of Unbiased Data

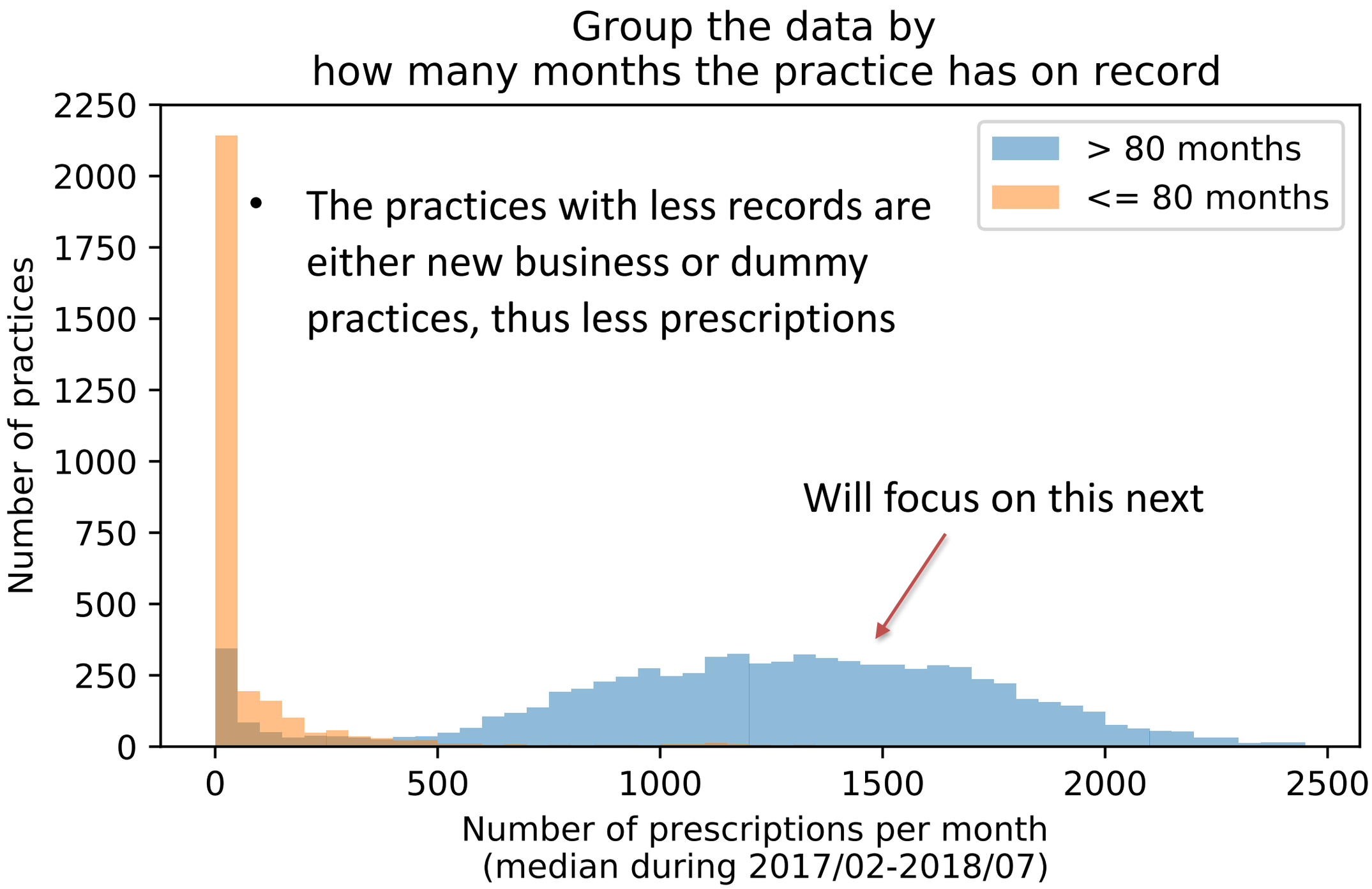


- Why 18-month window - long enough to be statistically significant and short enough so it does not catch long-term trends
- Clearly two components here - try to break it down

The Age of a Practice (How long it has been around)

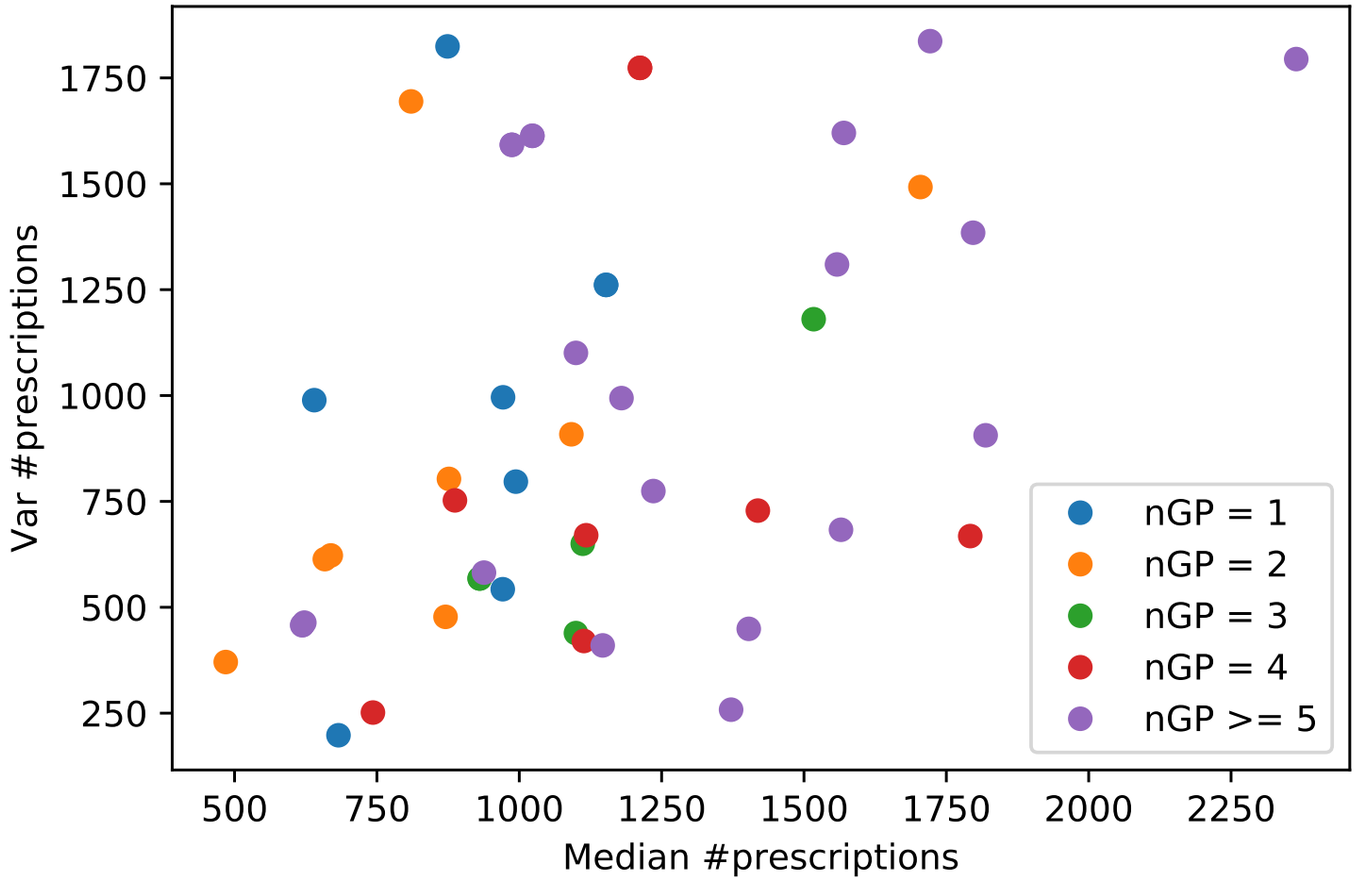


Performance by Age of Practice



#Practitioners per Practice

- Need to normalize performance by how many general practitioners(GP) each practice has.
- Randomly pick 50 practices and looked up their GPs.
- Assume distribution of practices with a given number of GPs is Gaussian, this shows the distributions for different number of GPs are heavily overlapping - clustering will be hard.
- Possible reason: records not up to date; some GPs are not full-time, etc. Will shoot for ‘effective' #GPs



K-means Clustering on the Variation

#prescriptions per month for 1 GP
= Poisson(k; \lambda) + Const.

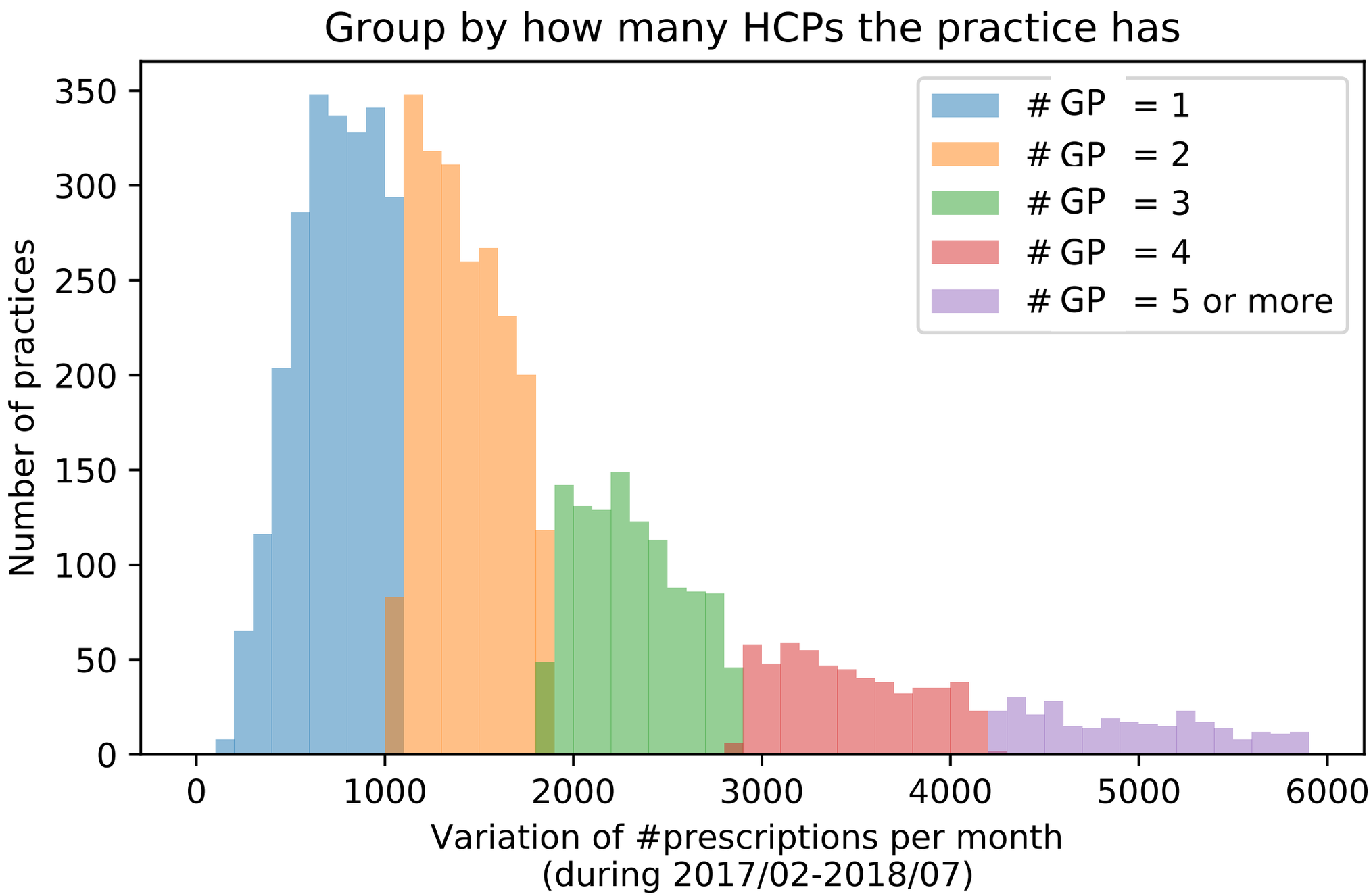
For incidental visits - accidents, flu, etc.

Monthly refills for patients with a chronological condition

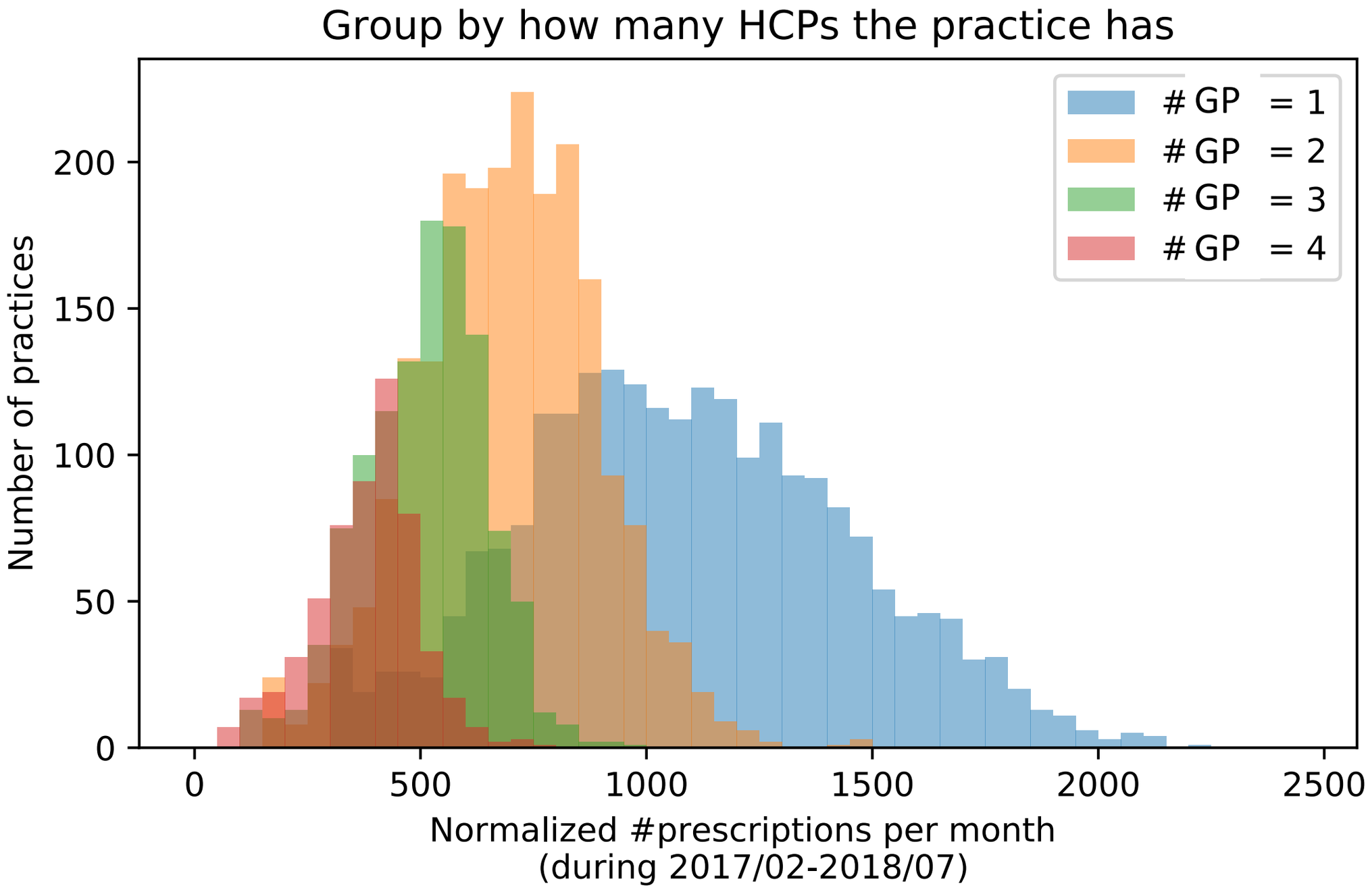
Var(#prescriptions for 1 GP) = \lambda

Var(#prescriptions for n GP) = n\lambda

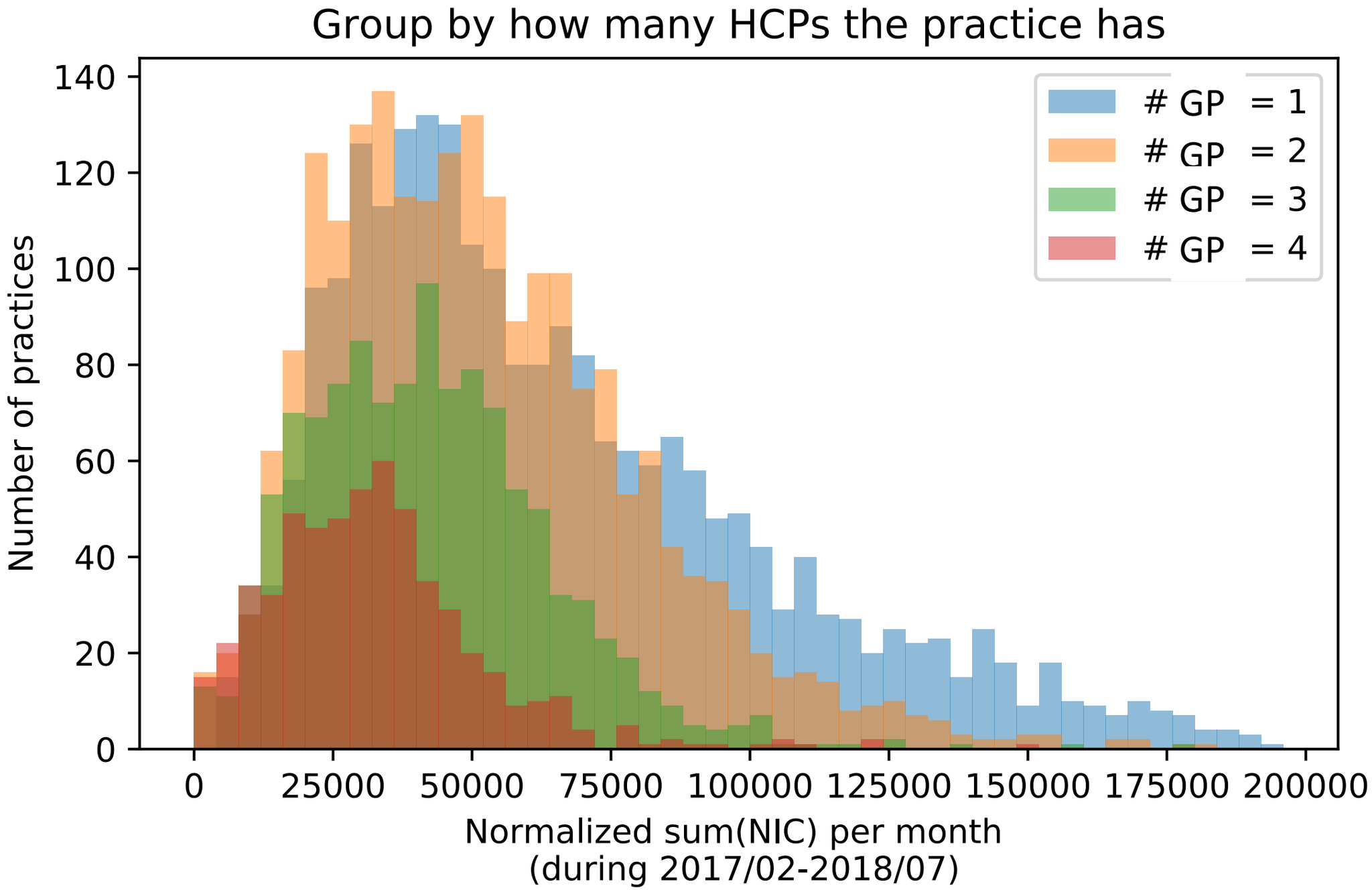
K-means Clustering on the Variation



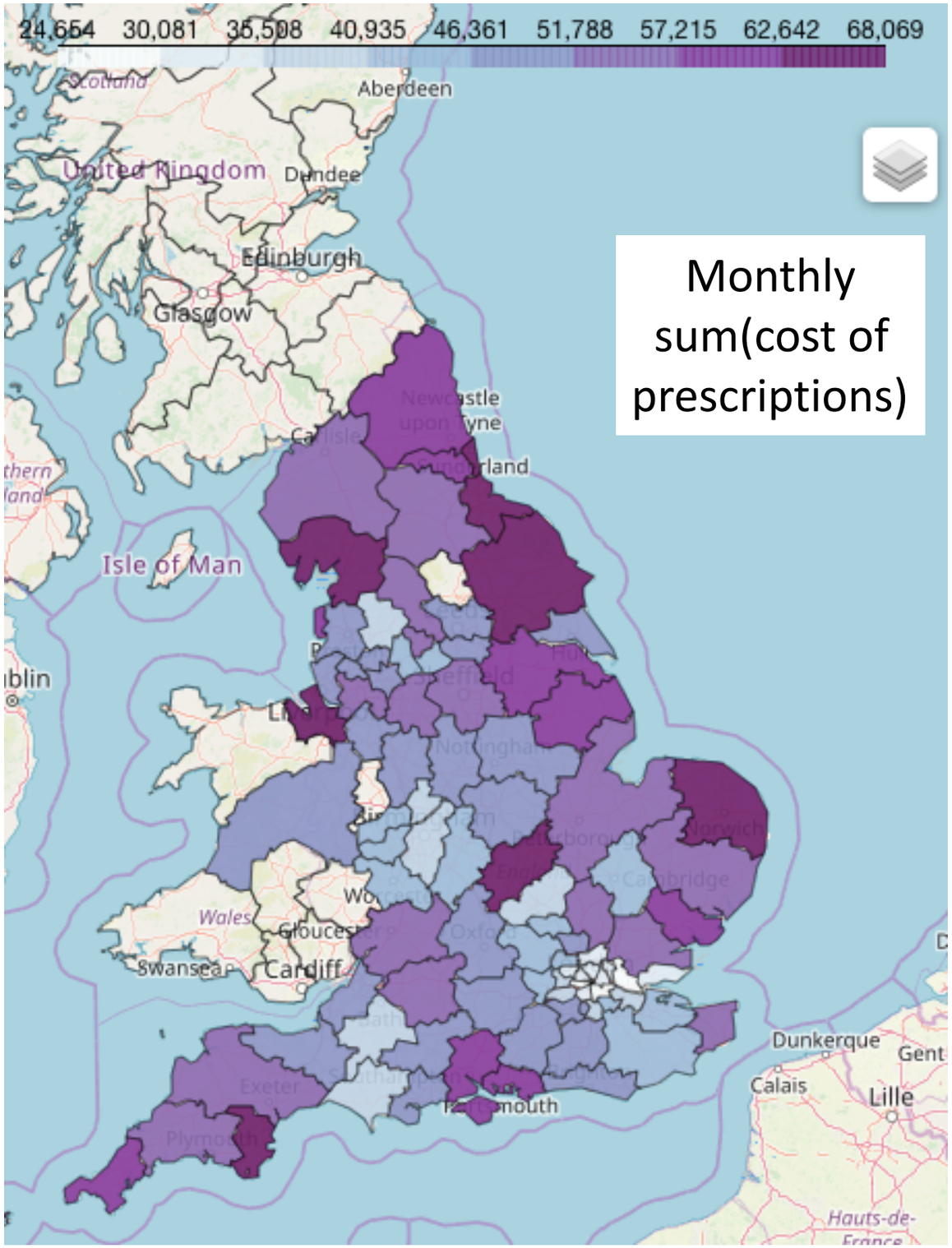
Performance Normalized after Clustering

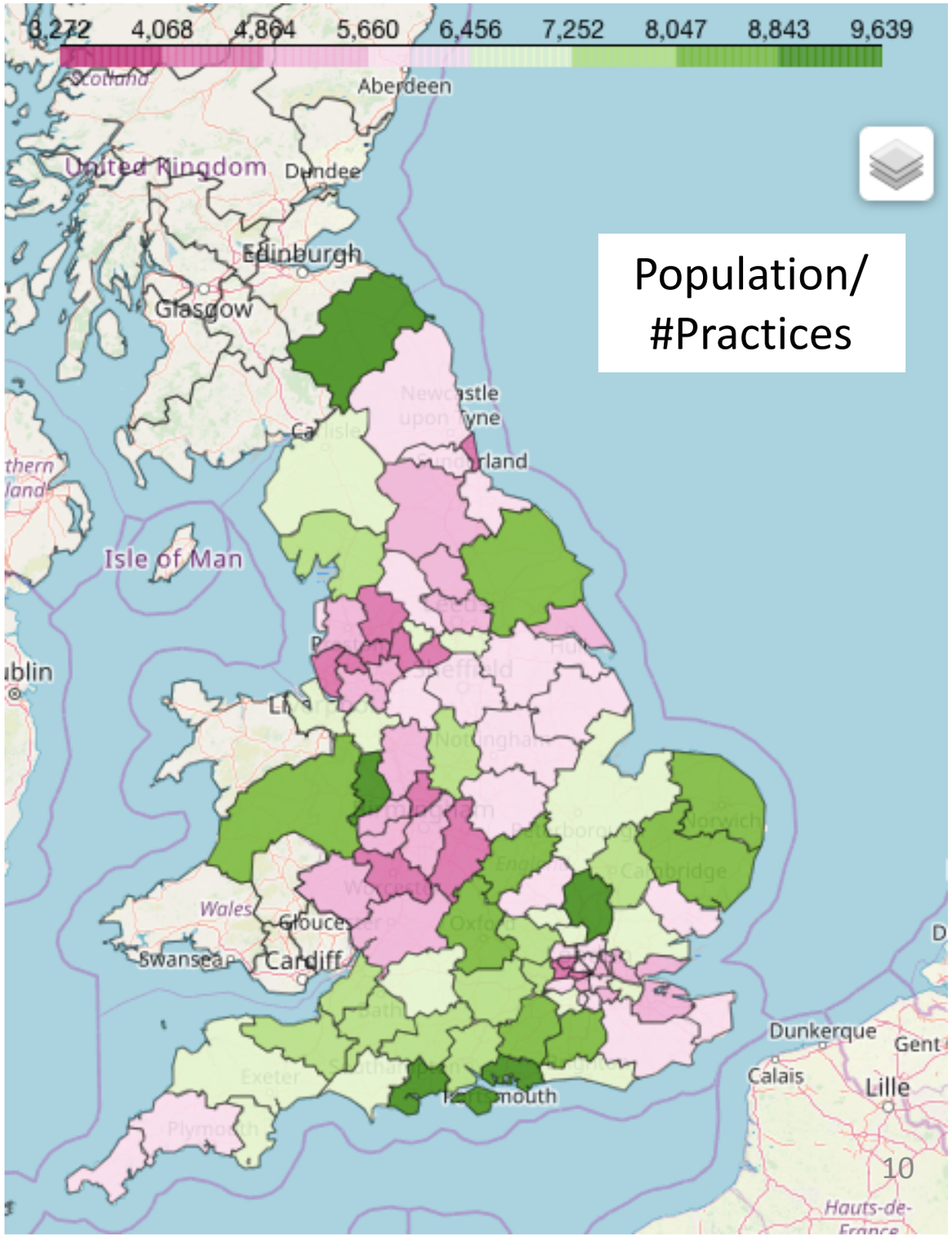


Performance Normalized after Clustering



Performance - Patient Practice Ratio





Conclusions

- Performance of a practice is evaluated by monthly number of prescriptions and monthly sum of cost of prescriptions, normalized by number of practitioners each practice has. Provides reference for marketing.

- Performance is largely determined by population/practitioner ratio. Regions with a large population/practitioner ratio warrants more attention from the marketing team.

Assumptions and Tools

- This data set is an accurate representation of the health practitioners in UK (A complete record / an unbiased sample)
- Practice code starts with an ‘Y’ are for dummy practices (and can be filtered out)

- Majority of the data (GB level) were processed by Spark. Analysis were done by SparkSQL and Python’s Pandas.

Data Source

- Prescription data on practice level - data.gov.uk/dataset/prescribing-by-gp-practice-presentation-level
- UK population by postcode area - www.nomisweb.co.uk
- Number of GPs in specific practices - www.nhs.uk