

Team 15: Featurization of long term CGM data to characterize menstrual cycle patterns in female patients with diabetes

Jane Li¹, Nathaniel Tillison¹, Jeyasri Venkatasubramani¹
¹Department of Bioengineering, University of California, San Diego
 Advisors: Dr. Benjamin Smarr, Saara Kriplani



Background

Diabetes Mellitus (DM): Metabolic disease where the body can't control blood glucose levels, with **589 million people worldwide** living with DM making it one of the greatest current health challenges¹

Estrogens and Glucose: Higher estrogens are associated with decreasing glucose levels and increased insulin sensitivity²

Long term biological rhythm sex differences in DM → contribute to personalized treatments for women based on menstrual cyclicity status

- Multiple studies with self-reported or CGM data have shown **phase based glucose changes**: increasing in ovulation, peaking in luteal phase, dropping steeply in menstruation^{3,4,5,6}
- Leverage **large cohort longitudinal CGM data** provided by Dexcom to investigate **prevalence and glucose features** with higher statistical power
- 2025's Team 29 discovered clear monthly cycles in daily CGM statistics – minutes out of range, percent out of range, & 95 percentile – in cyclical women

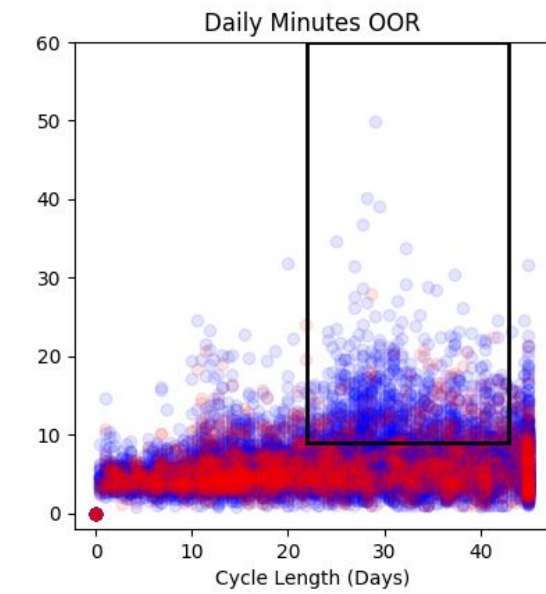
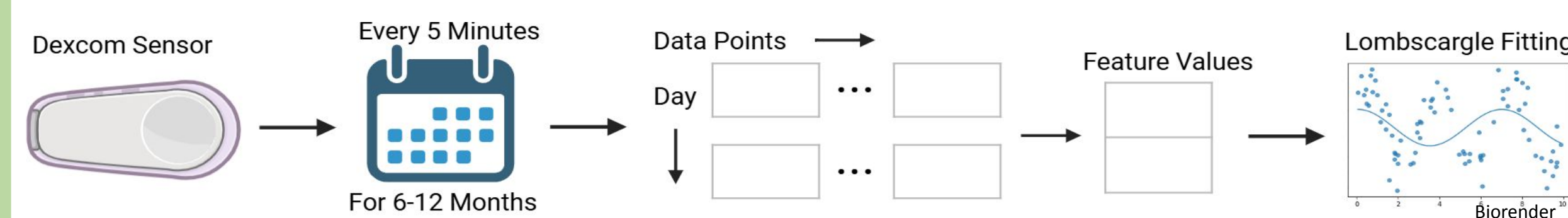


Figure 1: Lomb-Scargle of female (blue) vs. male (red) glucose metric

Objectives & Methods

Objective: Identify and analyze features displaying monthly cyclicity in women agnostically determined to be undergoing regular menstruation (women in 30s).



Dexcom sensors measure glucose every 5 minutes from diabetes patients for 6 months up to 1 year. Features are acquired from the resulting data and lombscargle is run on the data to model the cyclical patterns and strength.

Subproject 1 (Clean Data): Remove data abnormalities that could hinder the outcomes of the lombscargle pipeline.

Subproject 2 (Develop Multiscale Model Features): Test features that describe nuanced characteristics of rhythms: peak parametrization, circadian fluctuations, and within day cyclicity.

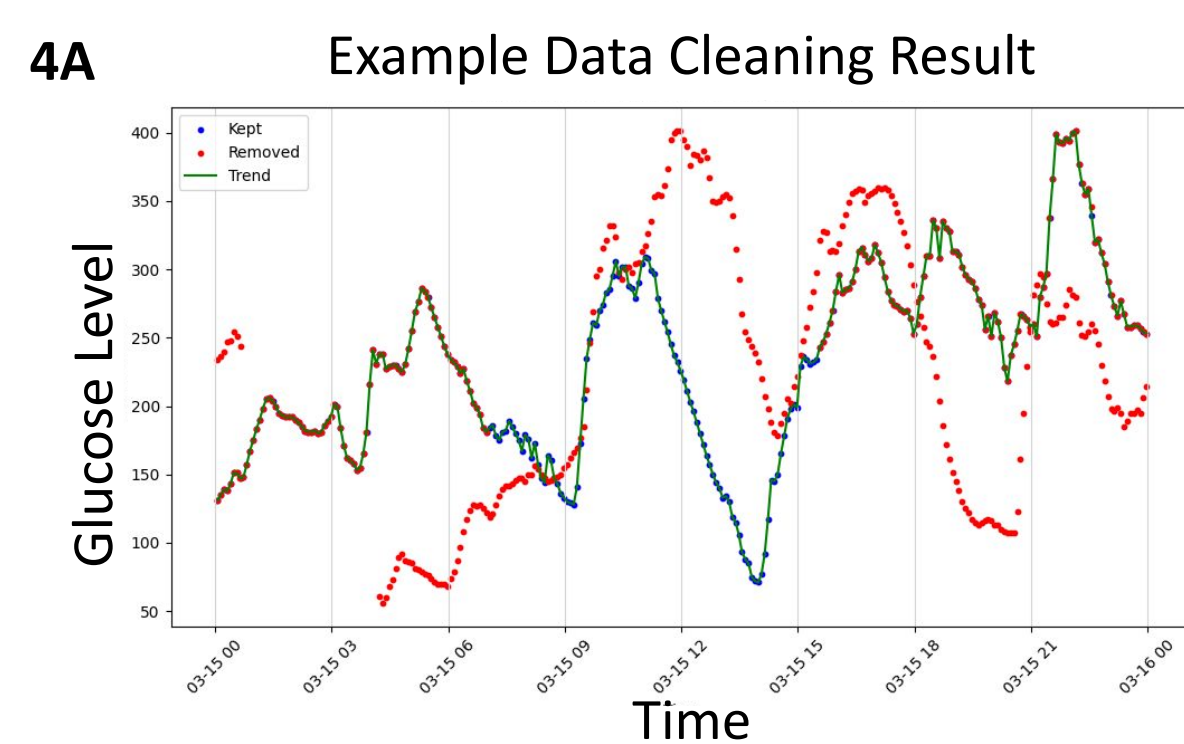
Subproject 3 (Feature Clustering and Stability): Characterize distinct groups of populations and identify features shared amongst individuals.

Cohort Table:

Sex	Age	Diabetes Type	Count
Female	30s	T1D	5000
		T2D	5000
	Total	10000	
	60s	T1D	1000
T2D		999	
Total	11999		
Male	30s	T1D	1000
		T2D	1000
	Total	2000	
	60s	T1D	1000
T2D		1000	
Total	4000		
Total	Total	15999	

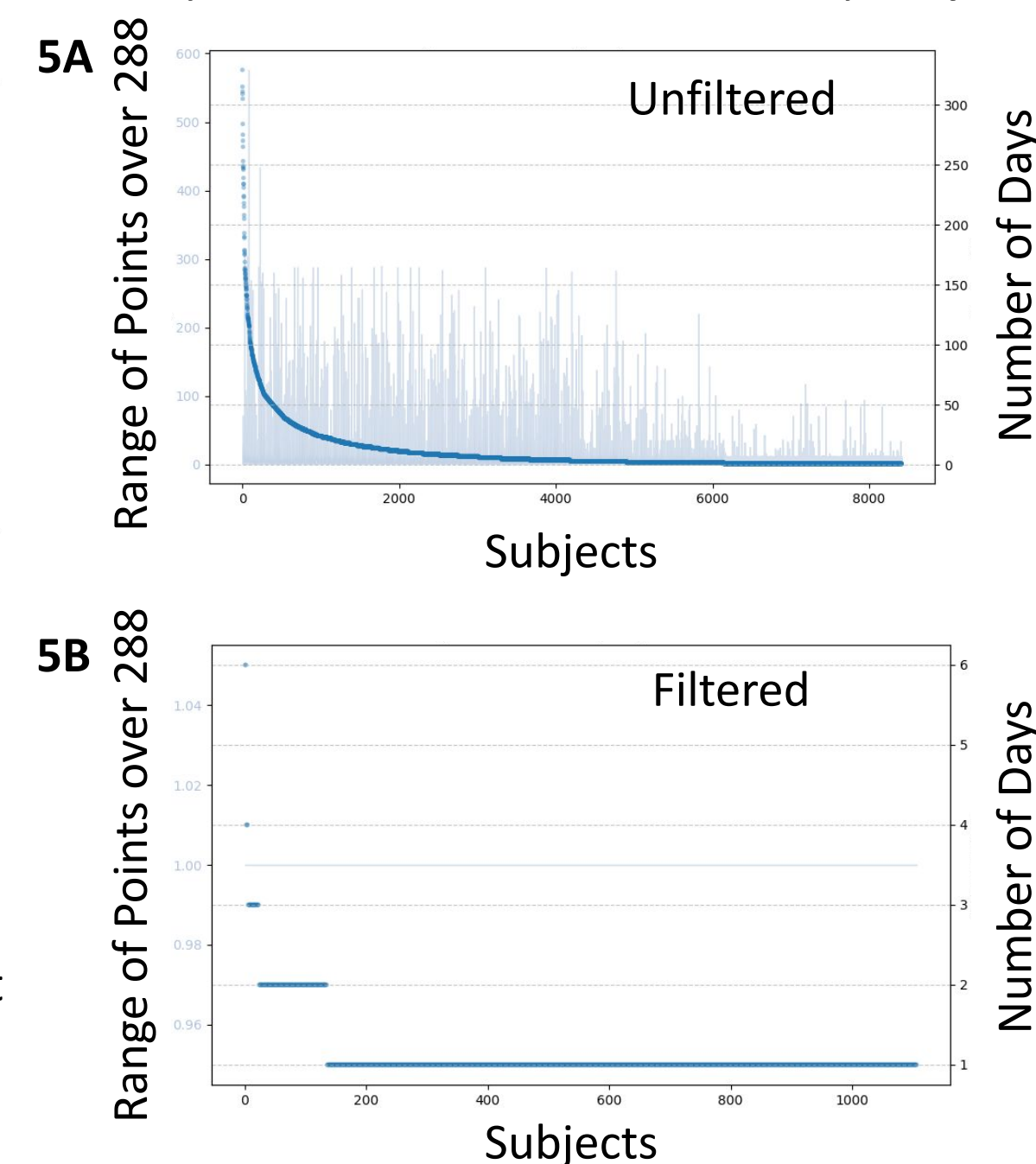
Results

Subproject 1: Remove Data Abnormalities



Total points removed: 1,063,073 (0.14%)
 Missing data interval (MDI):
 • 4.3% of subjects have different number of MDI - 0.42 + - 4.61 MDI
 • 4.4% of subjects have at least one different MDI length - 0.01 ± 0.26 minutes
 Subjects 6675 and 5231 were removed.

Days More Than 288 Data Points by Subject



Subproject 2: Test Multiscale Model Features

- Outlined box = 22-43 day peaks above a 8.94 power threshold = cyclic range

Legend: • Female • Male

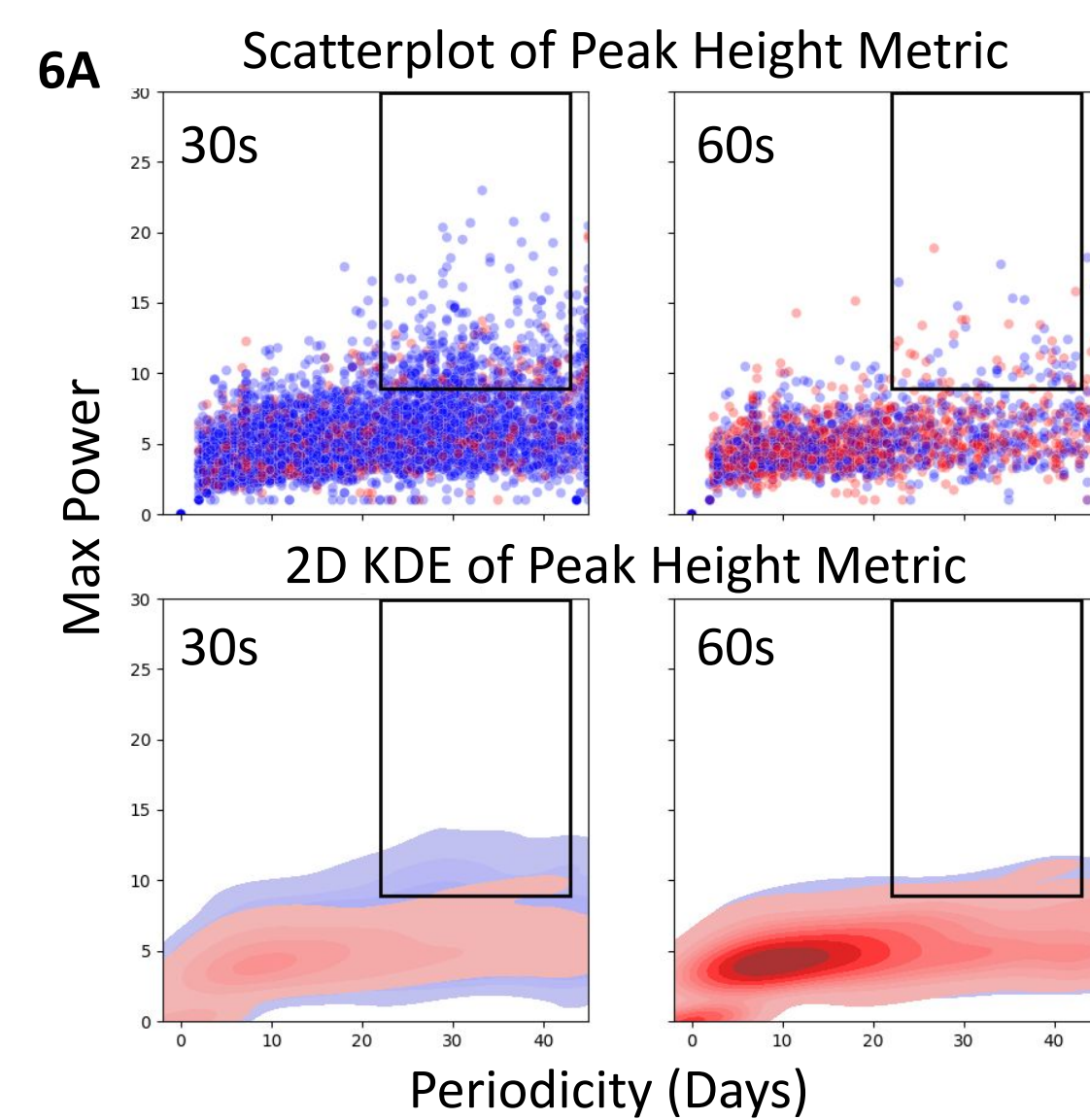


Figure 6A: Example Lomb-Scargle of female vs. male glucose metrics

KEY TAKEAWAY:

More females in 30s in the boxed cyclic range than other groups

Bold values = significant for p ≤ 0.05

Table 1: Chi-Square Test P-Values

Comparison	30s Males vs. 30s Females	60s Males vs. 60s Females
Peak Height	0.125	0.083
Peak Mid Height	0.003	0.823
Range	0.393	0.750
With in Day Cyclicity	0.918	0.655
Hourly Mean		
Hourly Variance		

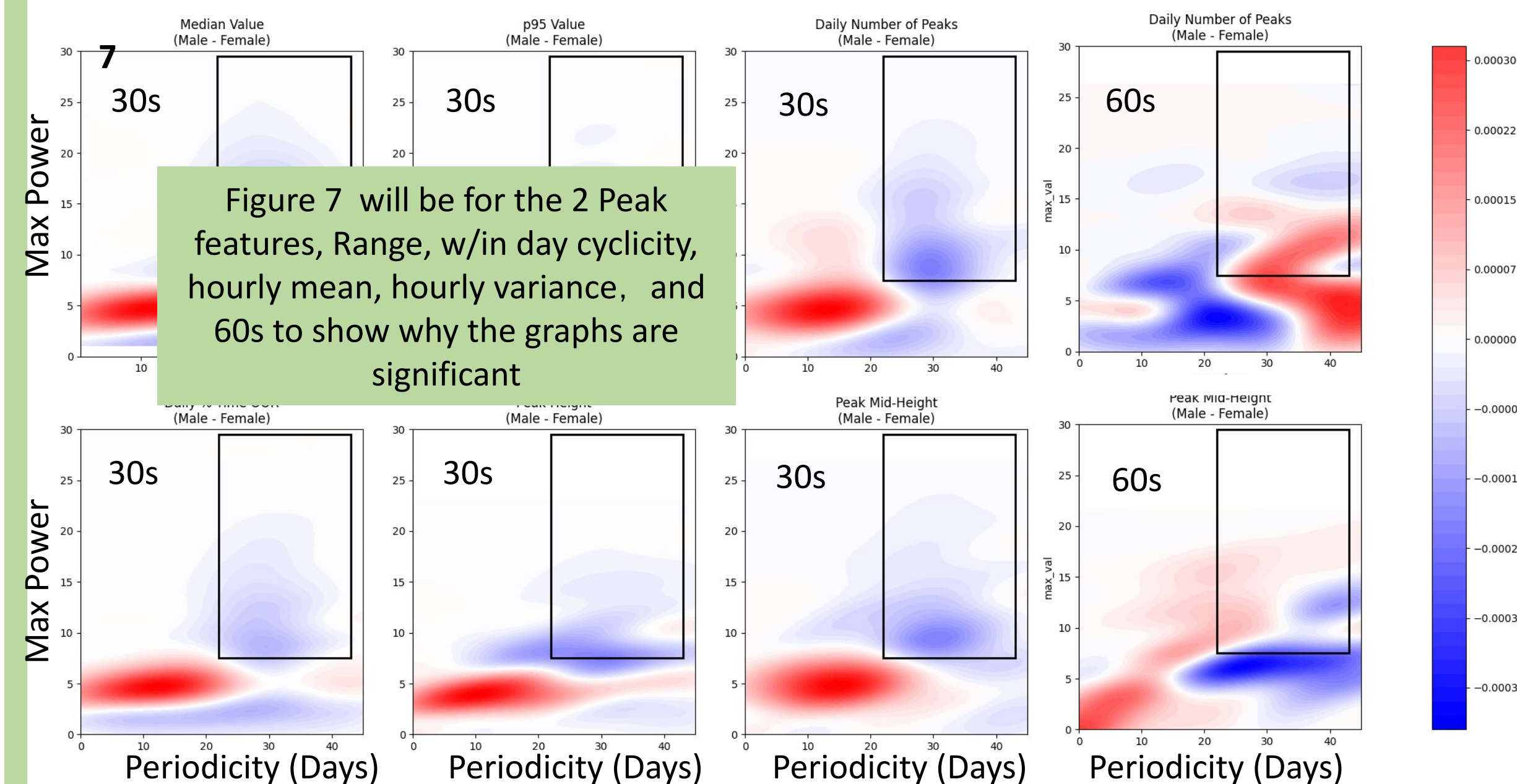


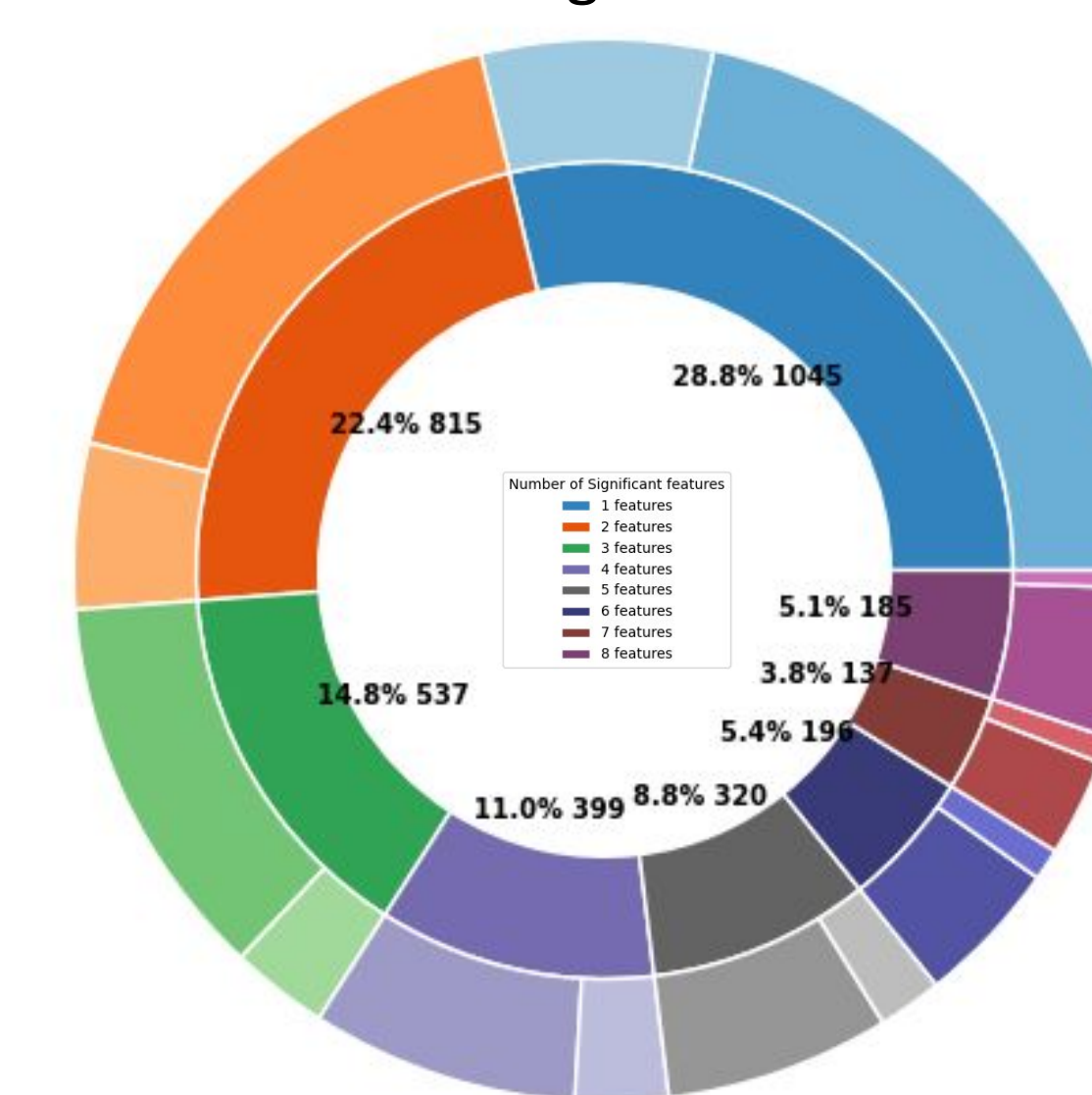
Figure 7 will be for the 2 Peak features, Range, w/in day cyclicity, hourly mean, hourly variance, and 60s to show why the graphs are significant

Figure 7: KDE of density differences (males - females)

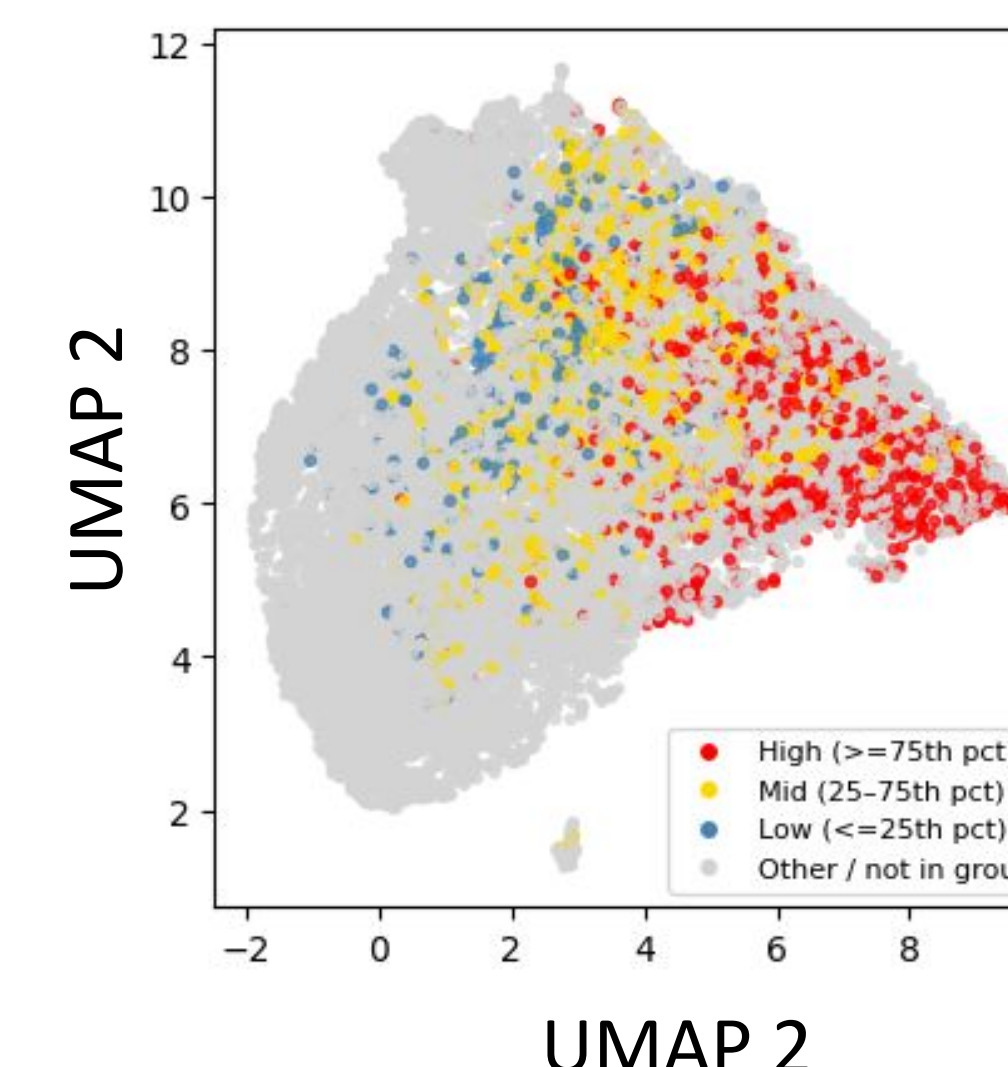
Results Cont.

Subproject 3: Feature Stability and Clustering

8b Percentage of Subjects with Different Amount of Significant Features



9 UMAP of Population Using Cyclical Features



Key Takeaway:

The results suggest that there are independent groups of people who show cyclicity in some features but not in others.

Conclusion & Future Directions

- There are **clear cyclic patterns in glucose data**, across a variety of features including peak height, peak mid-height, time OOR and % OOR which could potentially underscore differences in post prandial glucose regulation.
- There are variations in the features deemed commonly cyclic, with number of peaks and minutes OOR having the highest amount of overlap. Features pertaining to peaks and OOR could provide the most insight into menstrual cyclicity.
- Future directions include:
 - Exploring alternative algorithms and refine **Lomb-Scargle** parameters to reduce bias from z-normalization
 - Identifying cyclical glucose patterns aligned with menstrual phases using the mcPHASES dataset
 - Long term goal - contribute to **personalized DM treatment plans** based on an individual's cyclicity status.

Acknowledgements & References

We would like to thank our mentors Dr. Smarr and Saara Kriplani, as well as former Smarr Lab member Dr. Jamie Burks. We would also like to thank Dr. Taylor and the BENG 187 instructional team. Finally, thank you to Dexcom for providing the datasets for our project.



References:

Team 15: Featurization of long term CGM data to characterize menstrual cycle patterns in female patients with diabetes

Jane Li¹, Nathaniel Tillison¹, Jeyasri Venkatasubramani¹

¹Department of Bioengineering, University of California, San Diego

Advisors: Dr. Benjamin Smarr, Saara Kriplani



Background

Diabetes Mellitus (DM): Metabolic disease where blood glucose levels are abnormally regulated.

- 589 million adults worldwide¹
- Continuous Glucose Monitoring System (CGM) to monitor glucose levels continuously

Estrogens and Glucose: Higher estrogens are associated with decreasing glucose levels and increased insulin sensitivity³

Long term biological rhythm sex differences in DM

→ contribute to personalized treatments for women based on menstrual cyclicity status

- Multiple studies have shown **phase based glucose changes:** increasing in ovulation and luteal phase, dropping in late-follicular phase^{4,5,6}
- 2025's Team 29 discovered **clear monthly cycles in daily CGM statistics** – minutes out of range & 95 percentile – in cyclical women

Investigate prevalence and engineer multiscale model glucose features with large cohort longitudinal CGM data from Dexcom

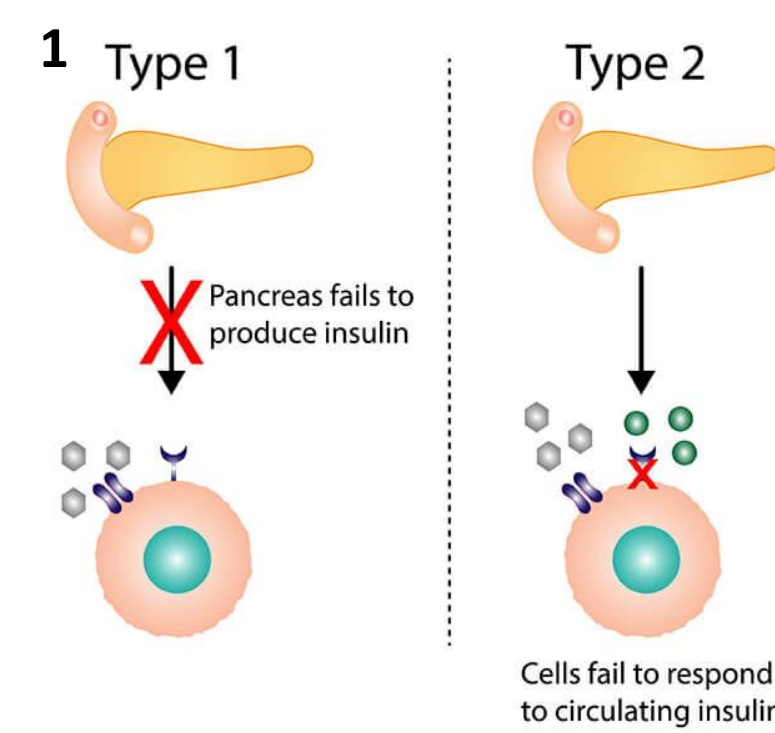


Figure 1: DM mechanism in DM 1 & DM 2²

Results

Subproject 1: Remove Data Abnormalities

2A Example Data Cleaning Result

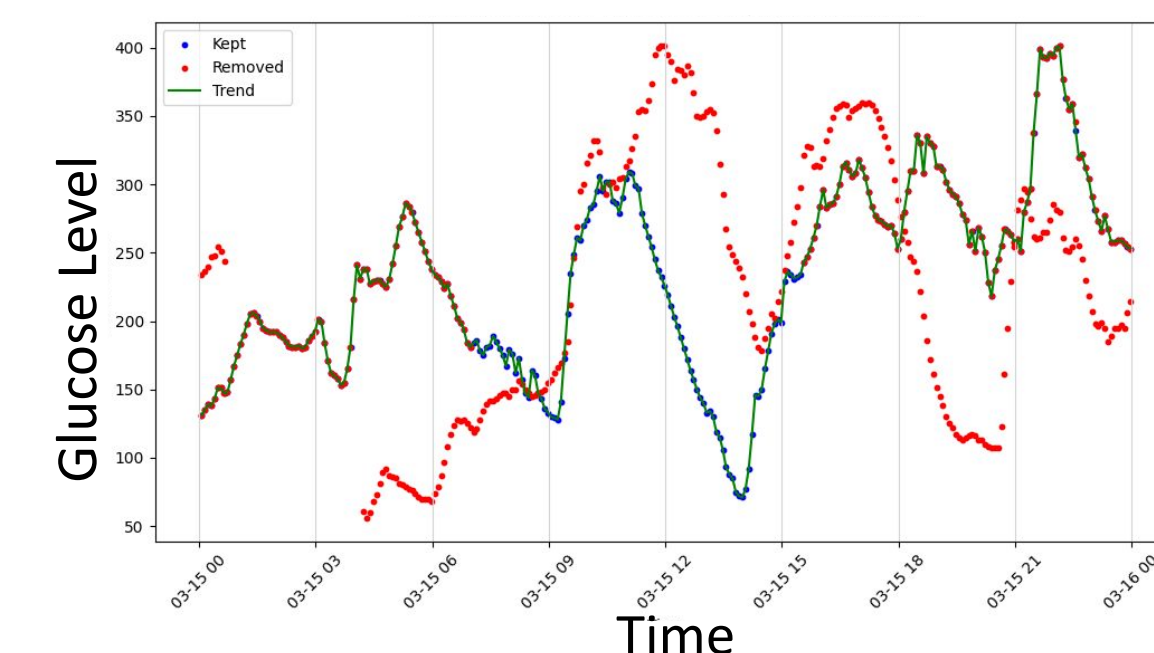


Figure 2A: Sample data cleaning result from one day of one subject.

- 0.14% of data removed
- Visually confirmed correct removal
- After filtering, only 1106 subjects have 1 datapoint over the limit compared to 8415 subjects before (Figure 2B)

2B

Days More Than 288 Data Points by Subject

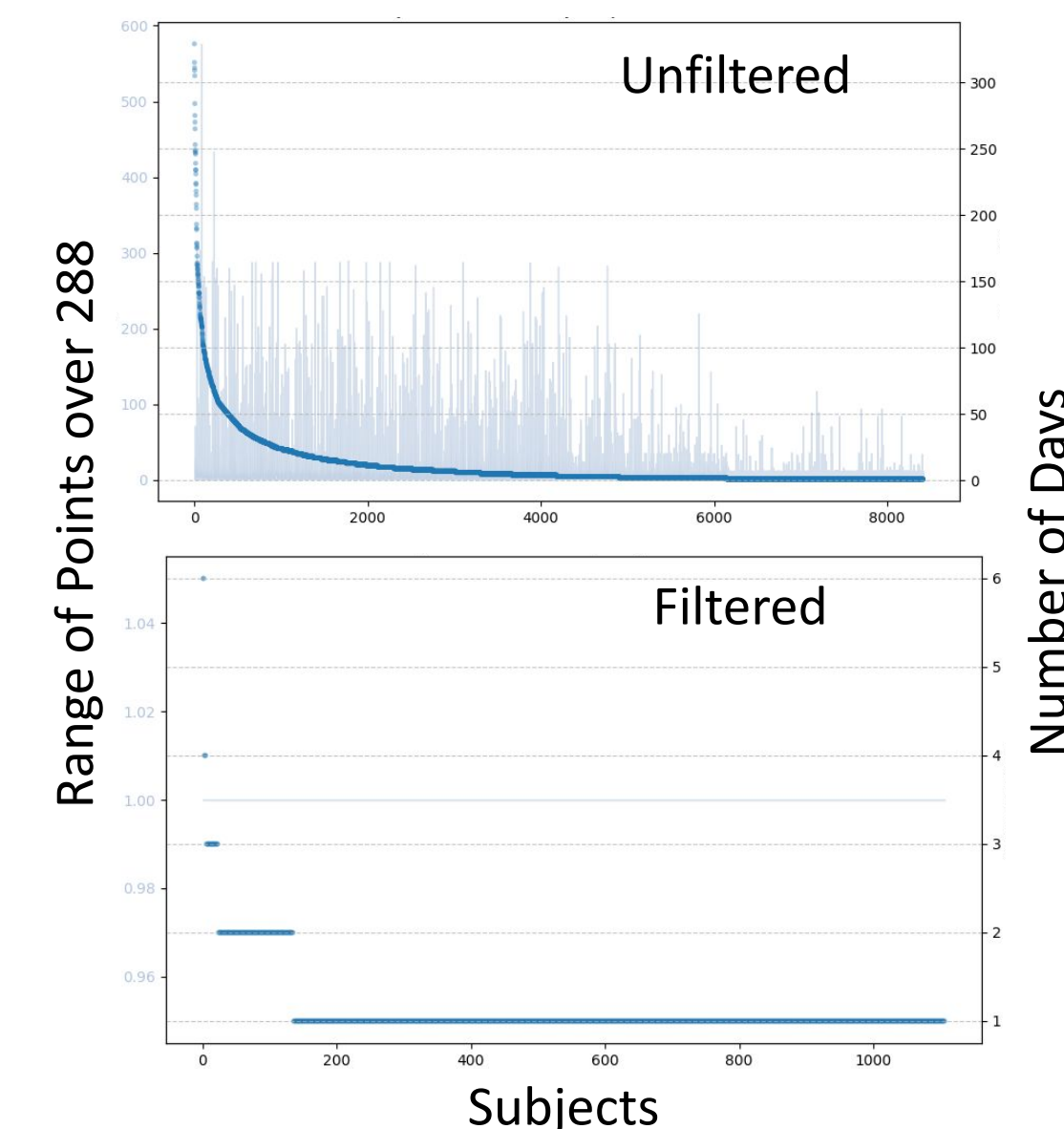


Figure 2B: Number of days and range over 288 data points of all subjects with at least one abnormal day before and after filtering

Subproject 2: Test Multiscale Model Features

3A Scatter Plot of Peak Height

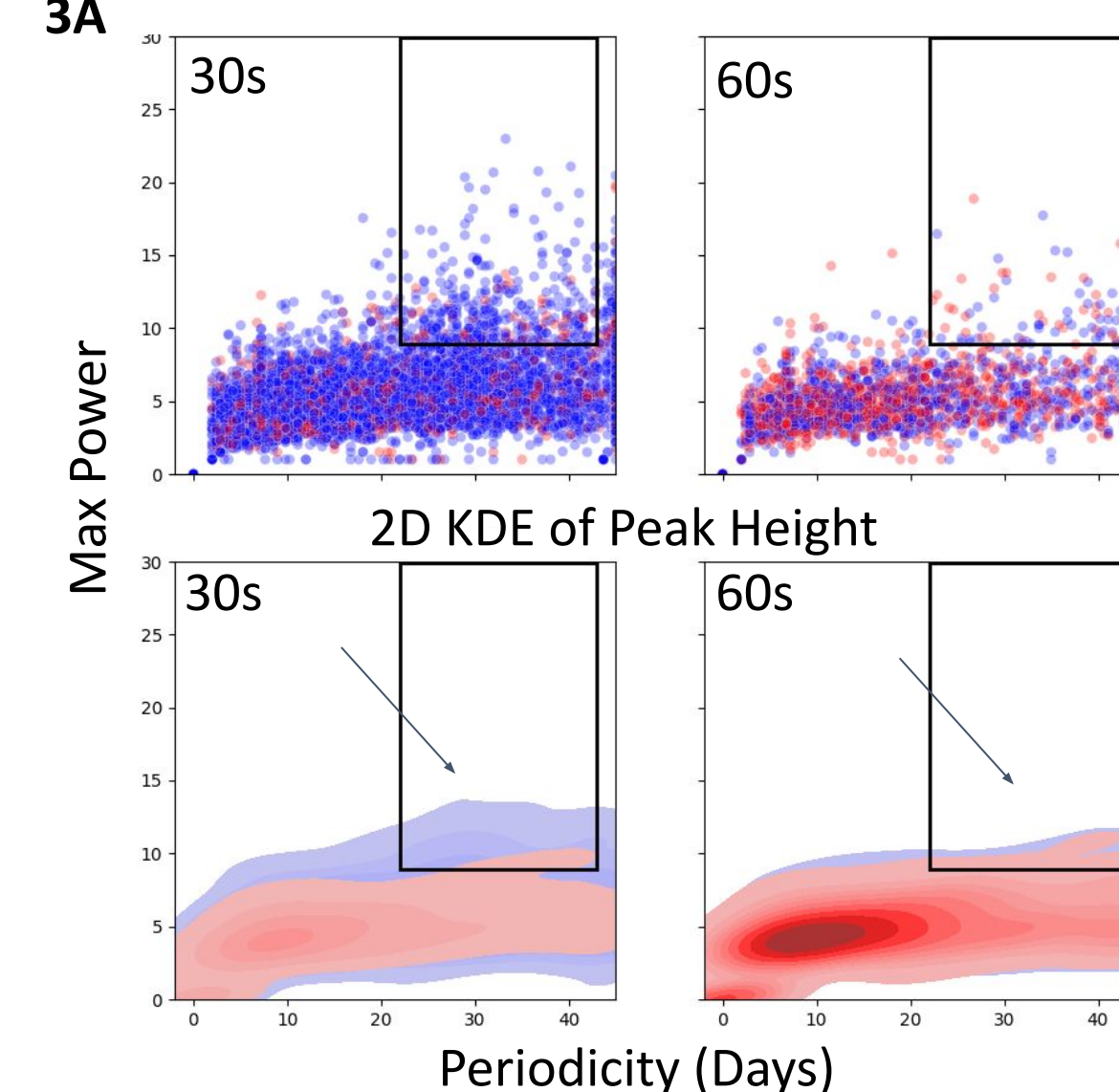


Figure 3A: Example Lomb-Scargle of female vs. male glucose metrics

3B Peak Mid Height

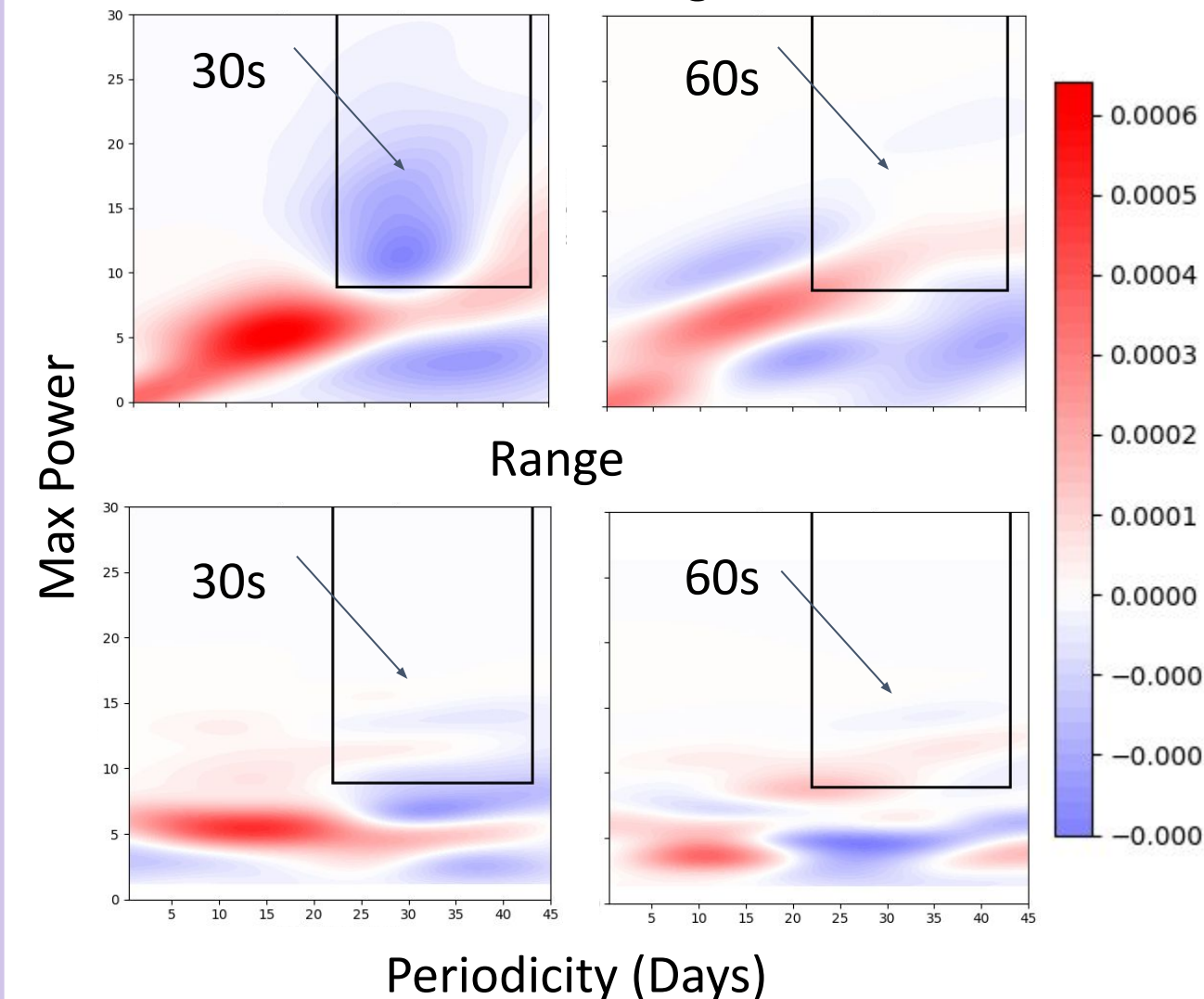


Figure 3B: KDE of density differences (males - females)

Table 1: Chi-Square Test P-Values

Comparison	30s Males vs. 30s Females	60s Males vs. 60s Females
Peak Height	1.63E-4	0.70
Peak Mid Height	3.35E-5	0.31
Peak Prominence	0.89	0.53
Range	0.32	0.83
Within Day Cyclicity	1	0.65
Circadian (Hourly Mean)	0.20	0.60
Circadian (Hourly Variance)	0.08	0.35

• Female • Male Significant for $p \leq 0.05$

Results Cont.

Subproject 3: Feature Stability and Clustering

4A UMAP of Population Using Cyclical Features

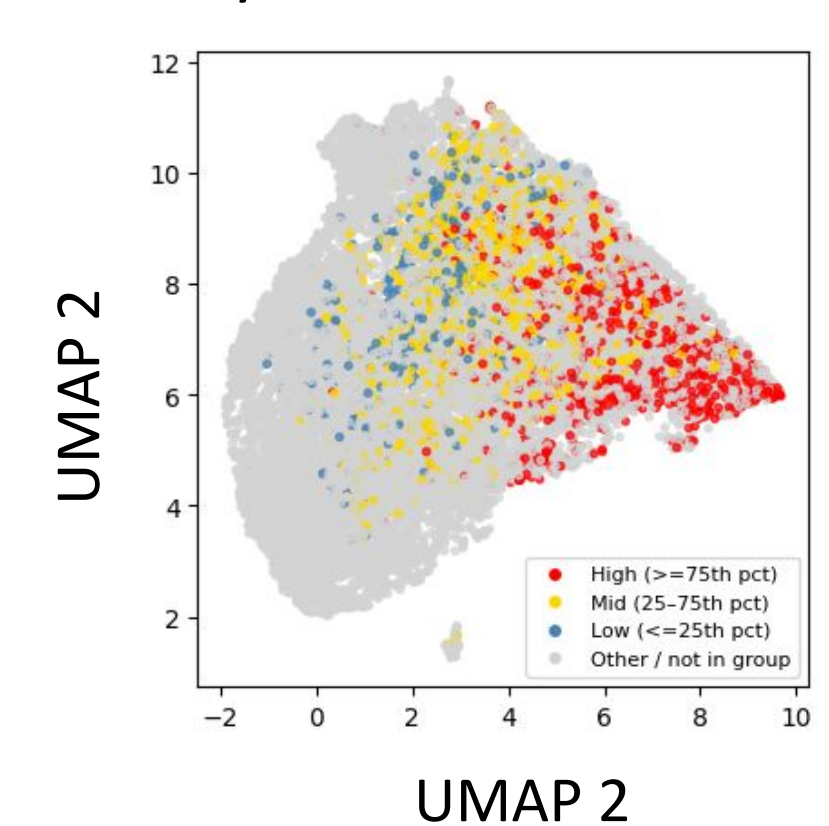


Figure 4A: Dimensional Reduction using UMAP from max power and period of the eight features.

4B Percentage of Subjects with Different Amount of Significant Features

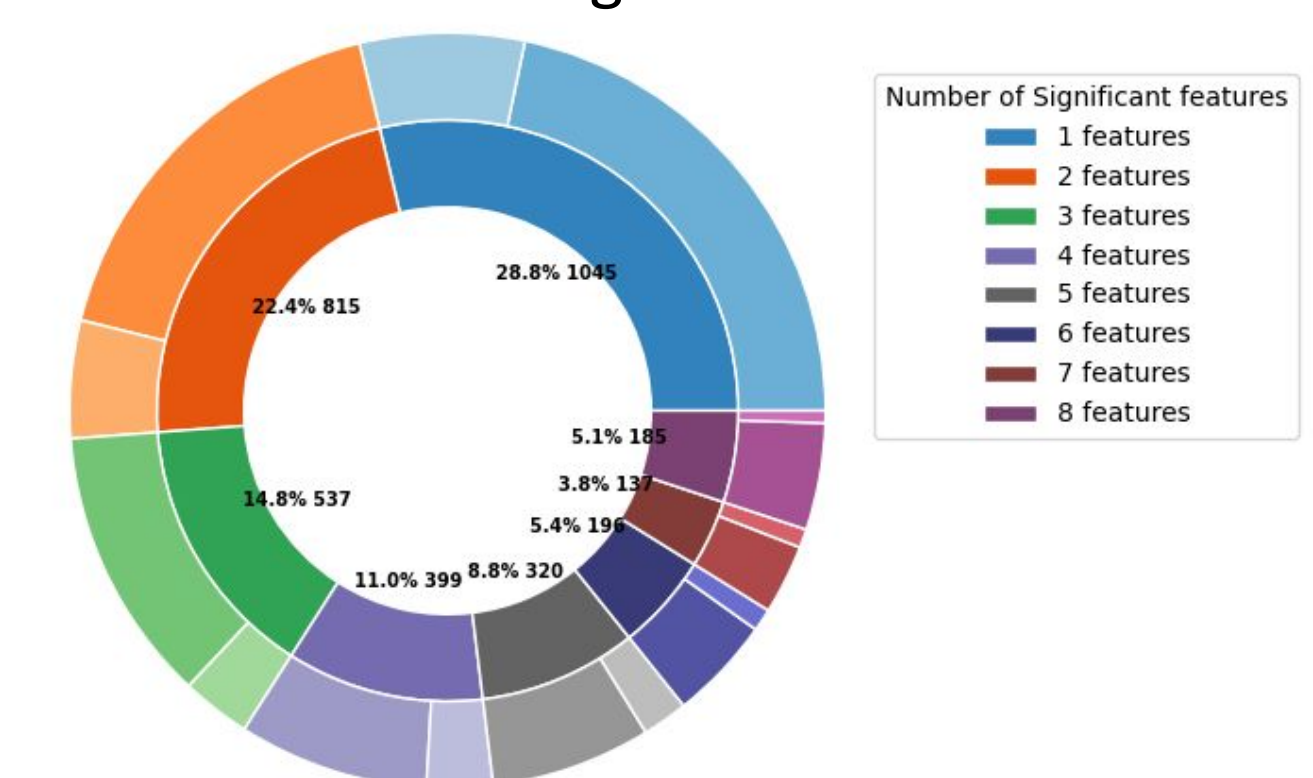


Figure 4B: Distribution of individuals with significance in X amount of features. Lighter color = males, darker color = females

- The UMAP reveals a continuous gradient, with samples arranged according to their level of cyclicity (Fig. 4A).
- The results suggest that there are independent groups of people who show cyclicity in some features but not in others (Fig. 4b).

Conclusion & Future Directions

- Certain features pertaining to **peaks**, which could characterize post meal time spikes, **had the most significance** in cyclicity. Relative height, which captures the height of the peak from the trough to crest, is more subject to change due to external factors. However, absolute height, a more simplistic measure, showed significance.
- **Simpler features capture cyclicity best.** It seems that complex features are heterogenous and are more so affected by external factors.
- Without **labels of those external factors**, it is difficult to extract true signal from complex features and validate our findings, which is a **major limitation** of the study.

Future directions:

- Reapplying dimensionality reduction after filtering by cyclicity range to better visualize groups with shared cyclic features.
- Identifying cyclical glucose patterns aligned with menstrual phases using the mcPHASES dataset.
- Long term goal - contribute to **personalized DM treatment plans** based on an individual's cyclicity status.

Acknowledgements & References

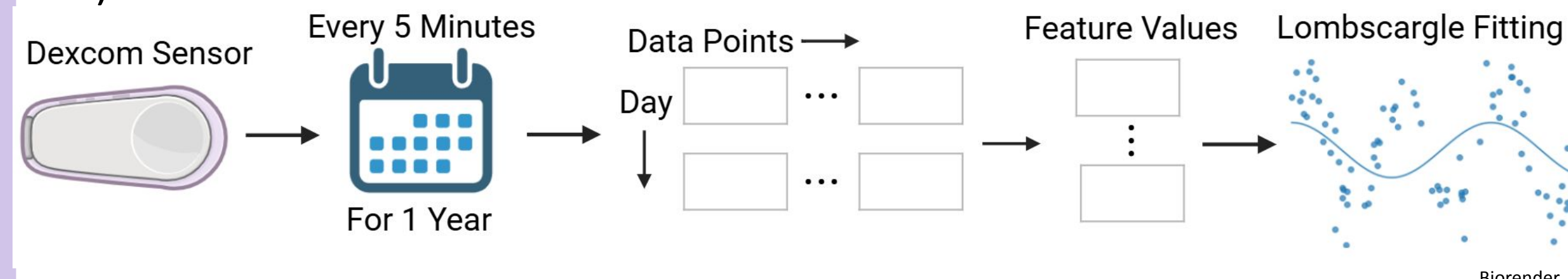
We would like to thank our mentors Dr. Smarr, Saara Kriplani and former Smarr Lab member Dr. Jamie Burks. We would also like to thank Dr. Taylor and the BENG 187 instructional team. Finally, thank you to Dexcom for providing the datasets for our project.



References:

Objectives & Methods

Objective: Identify and analyze features displaying monthly cyclicity in women agnostically determined to be undergoing regular menstruation (women in 30s).



Dexcom sensors measure glucose every 5 minutes from diabetes patients for 1 year. Features are acquired from the resulting data and Lomb-Scargle is run on the data to model the cyclical patterns and strength.

Subproject 1 (Clean Data): Remove data abnormalities that could hinder the outcomes of the lomb scargle pipeline.

Subproject 2 (Develop Multiscale Model Features): Test features that describe nuanced characteristics of rhythms: peak parametrization, circadian fluctuations, and within day cyclicity.

Subproject 3 (Feature Clustering and Stability): Characterize distinct groups of populations and identify features shared amongst individuals.

Cohort Table:

Sex	Age	Diabetes Type	Count
Female	30s	T1D	5000
		T2D	5000
		Total	10000
	60s	T1D	1000
		T2D	999
		Total	1999
Total		11999	
Male	30s	T1D	1000
		T2D	1000
		Total	2000
	60s	T1D	1000
		T2D	1000
		Total	2000
Total		4000	
Total		15999	

