Making Health Inferences from Wearable Device Models

Neal Cody, Isaac Oyediran, Justin Veerasami

[Dec 09, 2024] Healthcare Data Management Fall 2024

Physionet.org Dataset for Glycemic Variability and Wearable Devices

Major Data Formatting Distribution *(per patient)*

Primary:

- Datetime
  - (YYYY-MM-DD | 24H-TIME)
- Primary value
  - Singular variable measured for each parameter (glucose conc. Skin temp, etc.)

Secondary (*different formatting*):

- Accelerometer data (XYZ)
- Food Log Data
  - Many nutrition variables but all significant holistically

Monitoring Devices

Empatica E4 Wristband

Research oriented biometric wristband

Dexcom G6

Consumer oriented abdominal glucose monitor

Goals and Objectives

1. Comparing Glucose Responses to Meals Across Participants

Idea: Each group member selects one participant and tracks how their glucose levels fluctuate before and after meals. The project would involve comparing glucose trends for different participants and assessing whether certain individuals show stronger glucose responses after standardized meals.

Steps:

- Query glucose levels before and after meals using timestamps and food log data.
- Compare the magnitude and duration of glucose spikes after meals across participants.

Goal: Identify patterns in how participants' glucose levels respond to food intake

2. Analyzing the Impact of Physical Activity on Glucose Regulation

Idea: Each group member selects one participant and examines how their physical activity levels (based on accelerometry data) correlate with glucose stability. The goal is to determine if higher physical activity correlates with lower glucose variability or improved glucose control.

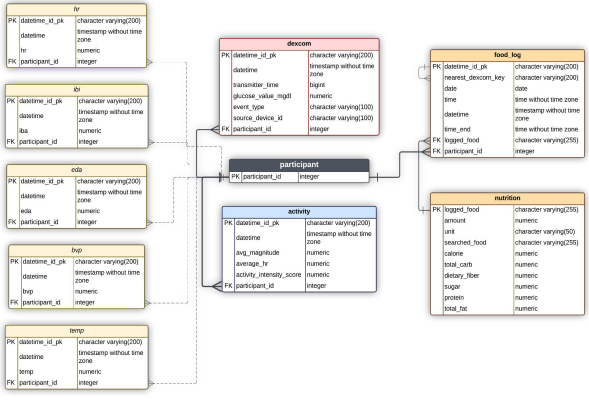
Steps:

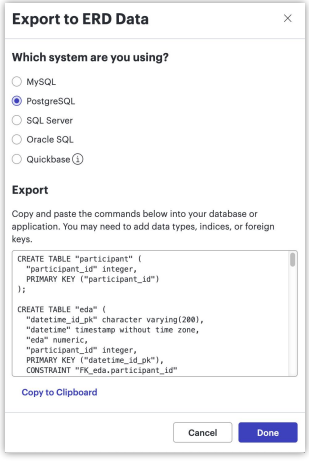
- Use accelerometry data to identify periods of high versus low physical activity.
- Track glucose levels during and after periods of activity.
- Compare the impact of physical activity on glucose stability across participants.

Goal: Understand whether increased physical activity helps participants maintain more stable glucose levels.

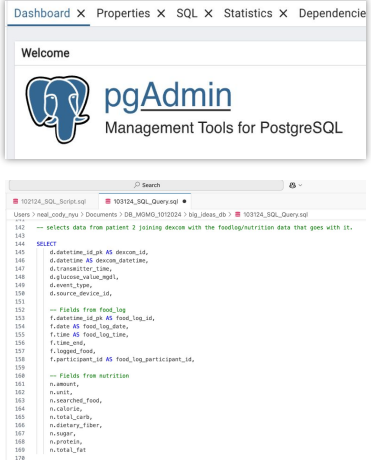
Reproducible Workflows: SQL dump file and useful functions enabled all team members to perform analysis/visualizations

LucidChart used for dynamic ERD coding





Build/Query Db



Reproducible Db / Code

Healthcare-db-mgmt-fall-2024 Project

The following Repository contains our project for Healthcare Database Management using openly-sourced data to generate a Database as a proof of concept to further understand the architecture behind designing relational databases

- This database primarily focuses on readings taken from activity data (derived from Empatica E4 accelerometry data) and food_log data to analyze readings taken by a Dexcom G6 glucose monitor

Link to the SQL Dump File (stored on google drive)

```
https://drive.google.com/file/d/1gMe2O7fD93vVu4l9xx7BBTWhw7wA_ajE/view?usp=sharing
```

How to Quickly Load the Database

In Postgres create a database to store the dump file within if you have not already

```
createdb -U [user] [database_name]
```

Run the PG Restore command and assign it as no owner to load the info without a role conflict

```
pg_restore -U [user] -d [database_name] --no-owner [filepath_to_SQLdump]
```

Useful Functions You Can Use Right Away

Get Participant Activity Data (example below)

- Allows you to obtain activity data associated with the selected participant
- Performs a join with the dexcom based glucose readings within the database to show the associated glucose readings

```
SELECT * FROM get_participant_activity_data(2);
```

Get Participant Food Log Data (example below)

- Allows you to obtain food log data associated with the selected participant
- Performs a join with the dexcom based glucose readings within the database to show the associated glucose readings
- The join is performed based off the nearest dexcom datetime as the foreign key as the food_log does not perfectly align with the glucose data but shows readings very close to similar times

```
SELECT * FROM get_participant_food_data(2);
```

Reference:

Cho, P., Kim, J., Bent, B., & Dunn, J. (2023). BIG IDEAs Lab Glycemic Variability and Wearable Device Data (version 1.1.2). PhysioNet. https://doi.org/10.13026/zthx-5212

Teams GitHub Page for Reproducible Analysis

Database_Connection_Test.Rmd	Add files via upload	last week
HealthcareDB Clean_archive.Rmd	fix: rename file	5 days ago
Healthcare_Db_Mgmt_R_viz_Final_12_9_24....	Add files via upload	3 hours ago
Healthcare_Db_Mgmt_R_viz_Final_12_9_24....	Add files via upload	3 hours ago
Healthcare_Db_Mgmt_R_viz_examples.Rmd	Add files via upload	2 weeks ago
Healthcare_Db_Mgmt_R_viz_examples.html	Add files via upload	2 weeks ago
README.md	Update README.md	6 hours ago

README

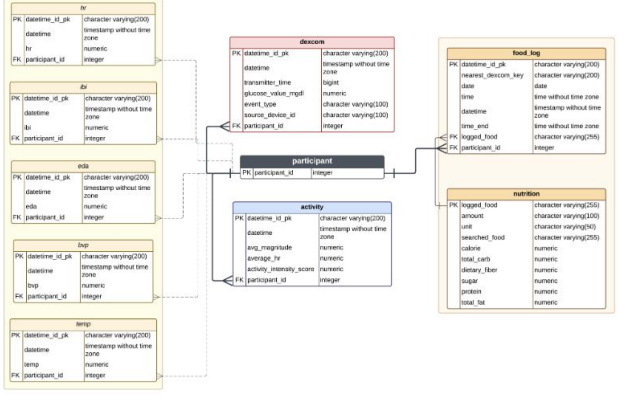
Healthcare-db-mgmt-fall-2024 Project

The following Repository contains our project for Healthcare Database Management using openly-sourced data to generate a Database as a proof of concept to further understand the architecture behind designing relational databases

- This database primarily focuses on readings taken from activity data (derived from Empatica E4 accelerometry data) and food_log data to analyze readings taken by a Dexcom G6 glucose monitor
- Other Tables are also included for the sake of cohesion of all of the data given through the website resources which are linked below

Please refer to the second branch of this repository for further detailed files used

ER Diagram of the Database



Link to the SQL Dump File (stored on google drive)

```
https://drive.google.com/file/d/1gMe2Q7fD93vVu4l9xx7BBTWhw7wA_ajE/view?usp=sharing
```

How to Quickly Load the Database

In Postgres create a database to store the dump file within if you have not already

```
createdb -U [user] [database_name]
```

Run the PG Restore command and assign it as no owner to load the info without a role conflict

```
pg_restore -U [user] -d [database_name] --no-owner [filepath_to_SQLdump]
```

Useful Functions You Can Use Right Away

Get Participant Activity Data (example below)

- Allows you to obtain activity data associated with the selected participant
- Performs a join with the dexcom based glucose readings within the database to show the associated glucose readings

```
SELECT * FROM get_participant_activity_data(2);
```

Get Participant Food Log Data (example below)

- Allows you to obtain food log data associated with the selected participant
- Performs a join with the dexcom based glucose readings within the database to show the associated glucose readings
- The join is performed based off the nearest dexcom datetime as the foreign key as the food_log does not perfectly align with the glucose data but shows readings very close to similar times

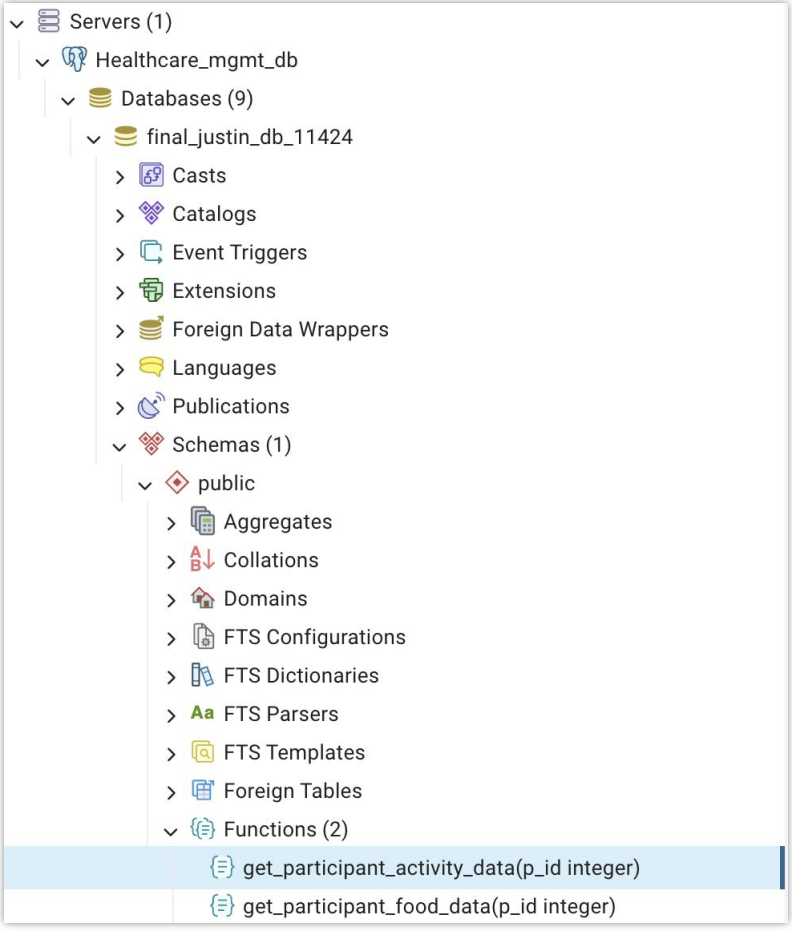
```
SELECT * FROM get_participant_food_data(2);
```

Reference:

Cho, P., Kim, J., Bent, B., & Dunn, J. (2023). BIG IDEAs Lab Glycemic Variability and Wearable Device Data (version 1.1.2). PhysioNet. https://doi.org/10.13026/zthx-5212

Two SQL functions streamlined queries and downstream analysis

pgAdmin4 - function feature



psql - unix command (Justin V)

```
CREATE OR REPLACE FUNCTION get_participant_activity
RETURNS TABLE (
    dexcom_id character varying,
    dexcom_datetime timestamp,
    transmitter_time bigint,
    glucose_value_mgdl numeric,
    event_type character varying,
    source_device_id character varying,
    activity_datetime timestamp,
    avg_magnitude numeric,
    average_hr numeric,
    activity_intensity_score numeric
) AS $$
BEGIN
    RETURN QUERY
    SELECT
        d.datetime_id_pk AS dexcom_id,
        d.datetime AS dexcom_datetime,
        d.transmitter_time,
        d.glucose_value_mgdl,
        d.event_type,
        d.source_device_id,
        a.datetime AS activity_datetime,
        a.avg_magnitude,
        a.average_hr,
        a.activity_intensity_score
    FROM dexcom d
    LEFT JOIN activity a ON d.datetime_id_pk = a.da
    WHERE d.participant_id = p_id;
END; $$
LANGUAGE plpgsql;
```

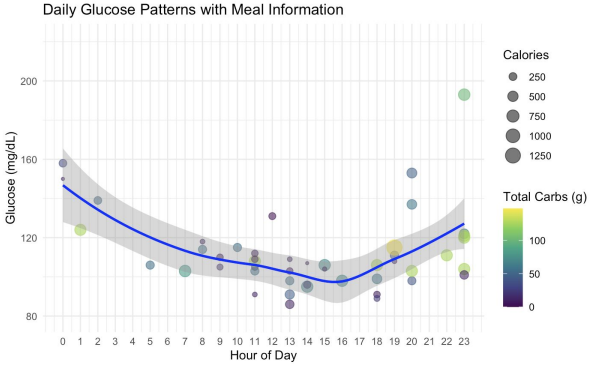
```
SELECT * FROM get_participant_activity_data(2);
```

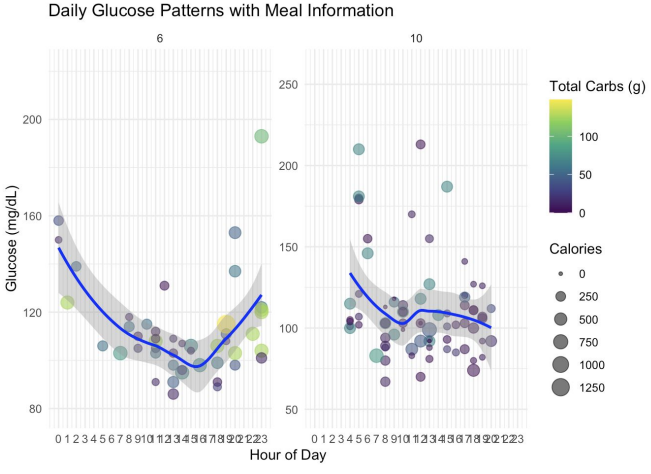
Rmd showcase

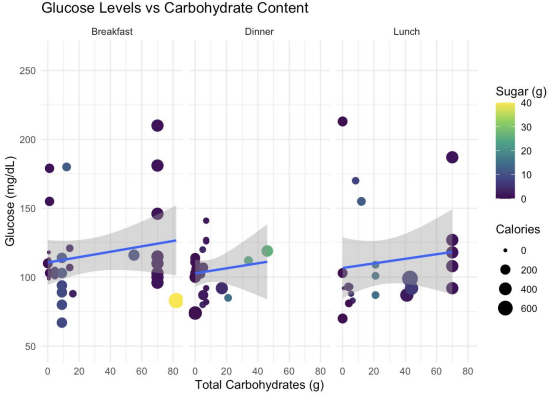
Can be referenced on our github after this presentation

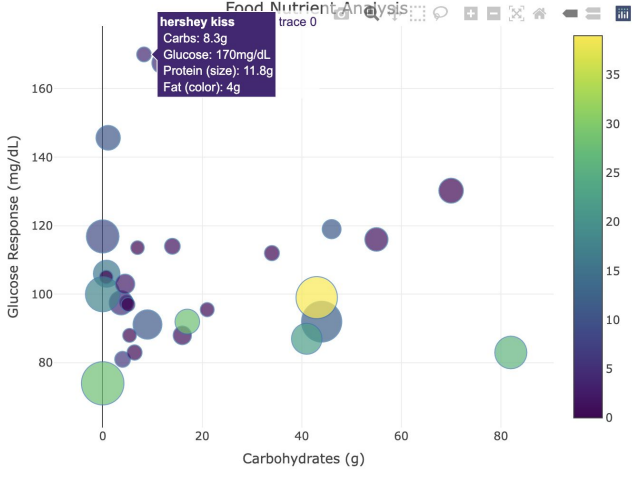
dbConnect() and dbGetQuery() dynamically query the PSQL database for visualization analysis

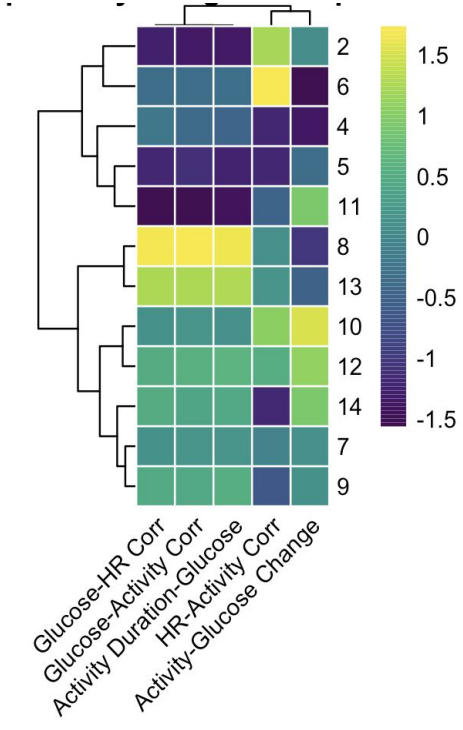
```r
# Replace with your actual database details
# this will return: <PqConnection> final_justin_db_11424@localhost:5432
con <- dbConnect(
  RPostgres::Postgres(),
  dbname = "final_justin_db_11424",   # Name of the database
  host = "localhost",                 # For local database
  port = 5432,                        # Default PostgreSQL port
  user = ,          # Your PostgreSQL username
  password =       # Your PostgreSQL password
)
```







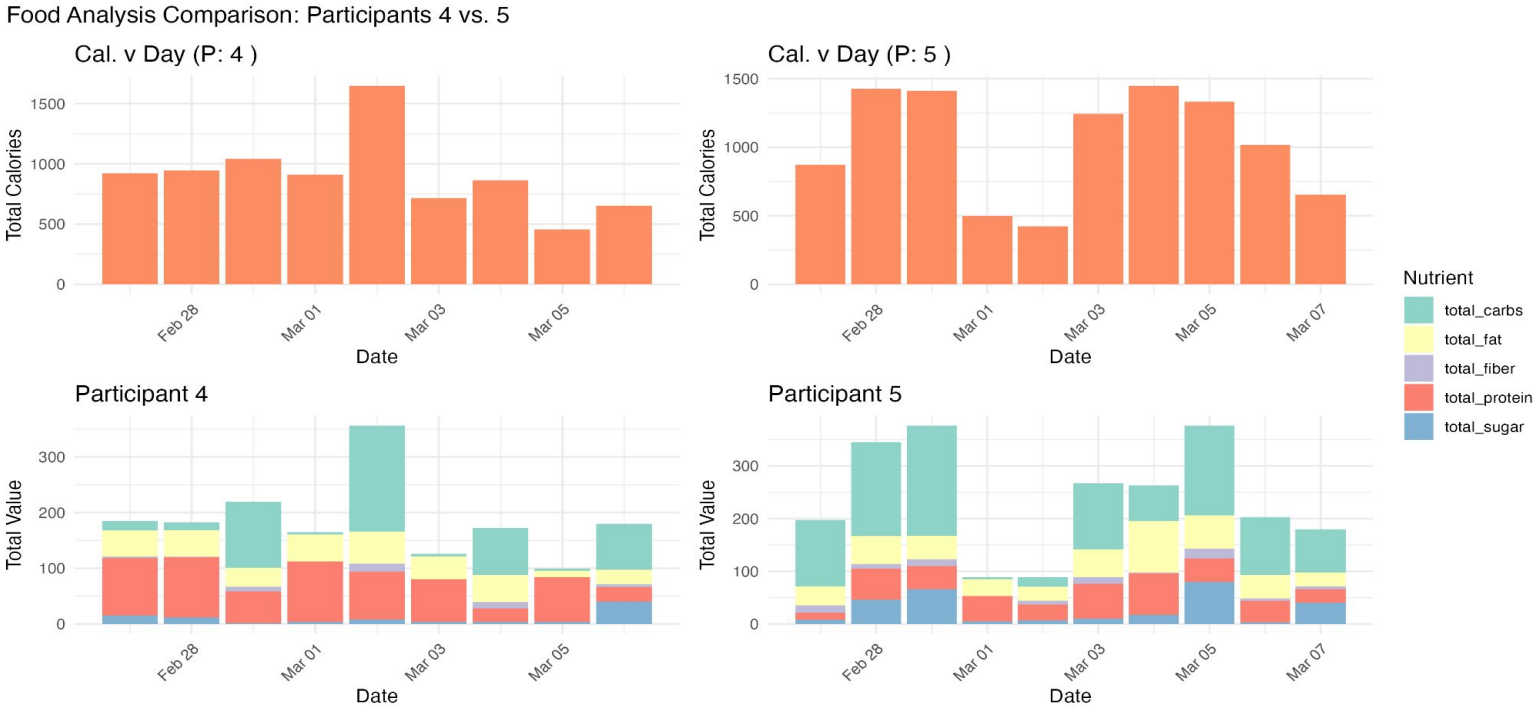




Participant #	Gender	HbA1c
4	Female	6.4%
5	Female	5.7%

HbA1c Ranges		
Normal	Prediabetes	Diabetes
< 5.7%	5.7% - 6.4%	>6.5%

Dietary Patterns: Calories Over Time



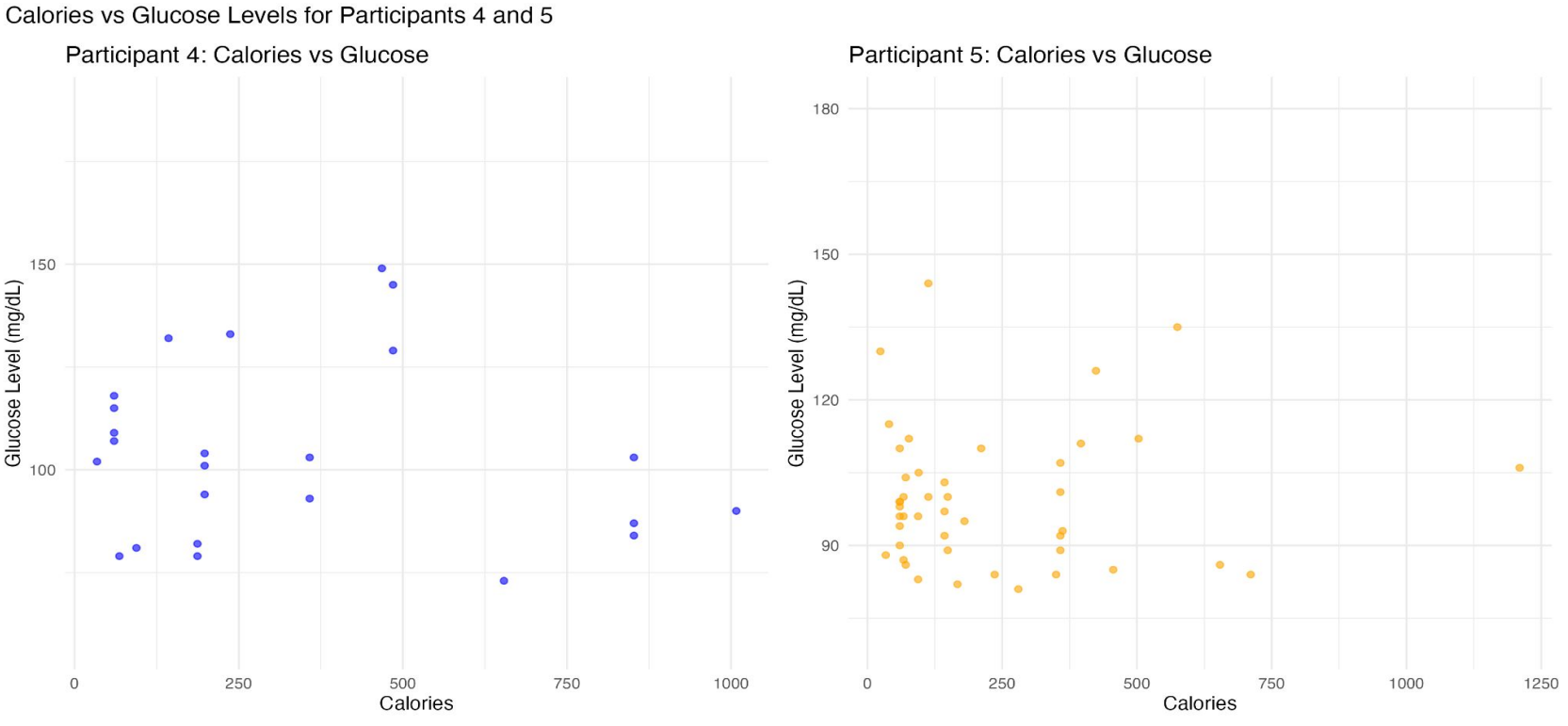
Key Takeaways:

- Participants 4 and 5 show distinct dietary patterns in terms of daily calorie intake and nutrient distribution
- Participant 5 consistently consumes more calories across days, the nutrient distribution (carbs, fats, proteins) remains relatively similar between participants

Participant #	Gender	HbA1c
4	Female	6.4%
5	Female	5.7%

HbA1c Ranges		
Normal	**Prediabetes**	**Diabetes**
< 5.7%	5.7% - 6.4%	>6.5%

Food Impact On Glucose Levels



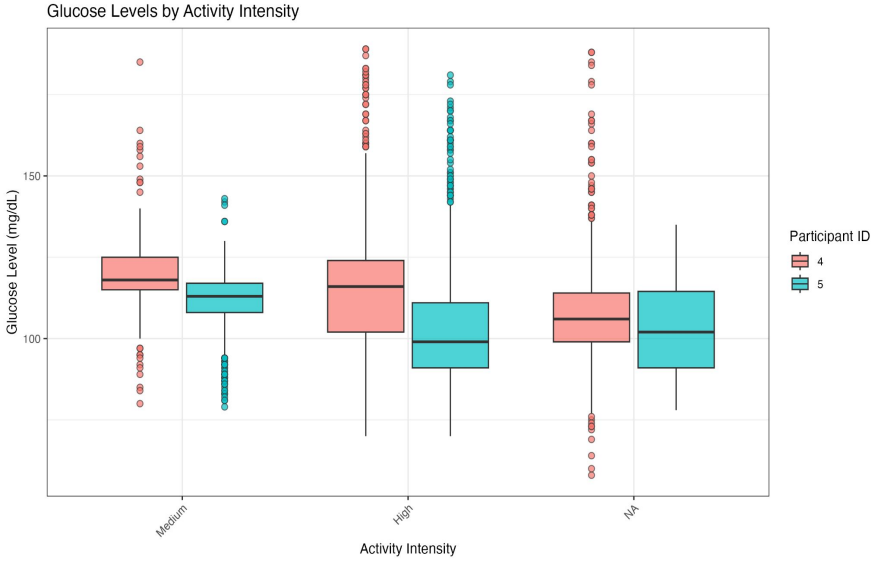
Key Takeaways:

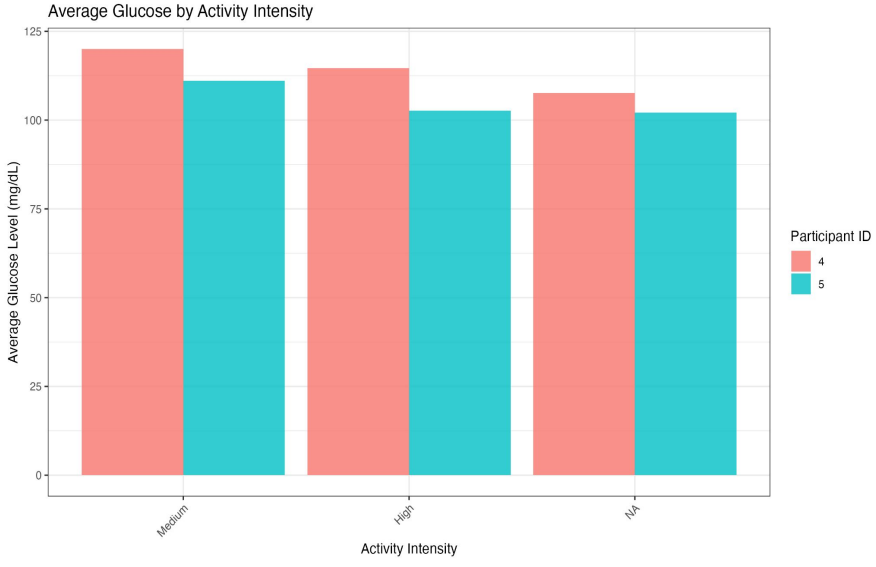
- Participant 4 shows fewer meals with higher calorie content, but these meals correspond to larger glucose variability
- Participant 5 has more meals logged, with glucose levels showing less variability across different calorie intakes, especially at lower calorie levels

Participant #	Gender	HbA1c
4	Female	6.4%
5	Female	5.7%

HbA1c Ranges		
Normal	**Prediabetes**	**Diabetes**
< 5.7%	5.7% - 6.4%	>6.5%

Activity Impact on Glucose Levels





Key Takeaway:

- Glucose levels tend to decrease as activity intensity increases for both participants, with high-intensity activity showing narrower variability compared to medium activity

Key Takeaway:

- On average, both participants show similar glucose levels across activity intensities

References and Acknowledgments

Cho, P., Kim, J., Bent, B., & Dunn, J. (2023). **BIG IDEAs Lab Glycemic Variability and Wearable Device Data** (version 1.1.2). PhysioNet. https://doi.org/10.13026/zthx-5212

Bent, B., Cho, P.J., Henriquez, M. et al. **Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches.** npj Digit. Med. 4, 89 (2021). https://doi.org/10.1038/s41746-021-00465-w

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). **PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals.** Circulation [Online]. 101 (23), pp. e215–e220.

GitRepo Link - https://github.com/NCodyBmi2024/Healthcare-db-mgmt-fall-2024

Acknowledgments: Jacob Swanson, Jacob Stimes, and Mark Grivanis

Appendix

Project Plan

Mid September - October 16

Phase 1: Project Proposal & Setup

Selected our dataset, outlining project goals, and defining research questions

Key activities

- Selected dataset and designed data architecture
- Conducted literature review and formulated research questions
- Established data storage location

October 17 - November 4

Phase 2: Data Cleaning & Structure

Prepare data for analysis through cleaning and structuring with an ER diagram

Key activities

- Clean and consolidate all data for consistency and accuracy
- Create ER diagram to support efficient querying and data visualization
- Validate data accessibility and usability for all team members

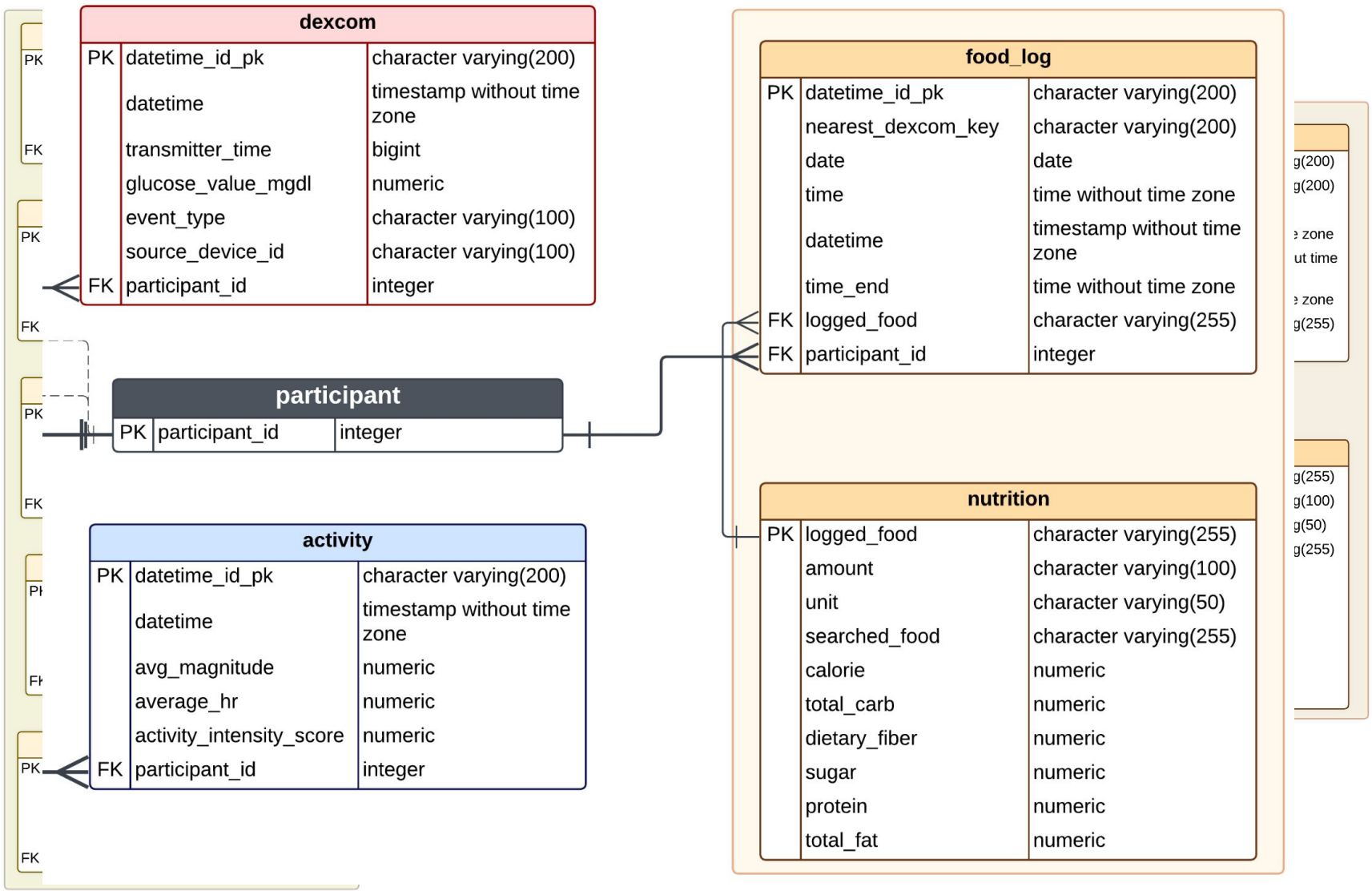
November 5 - December 9

Phase 3: Answering Research Questions

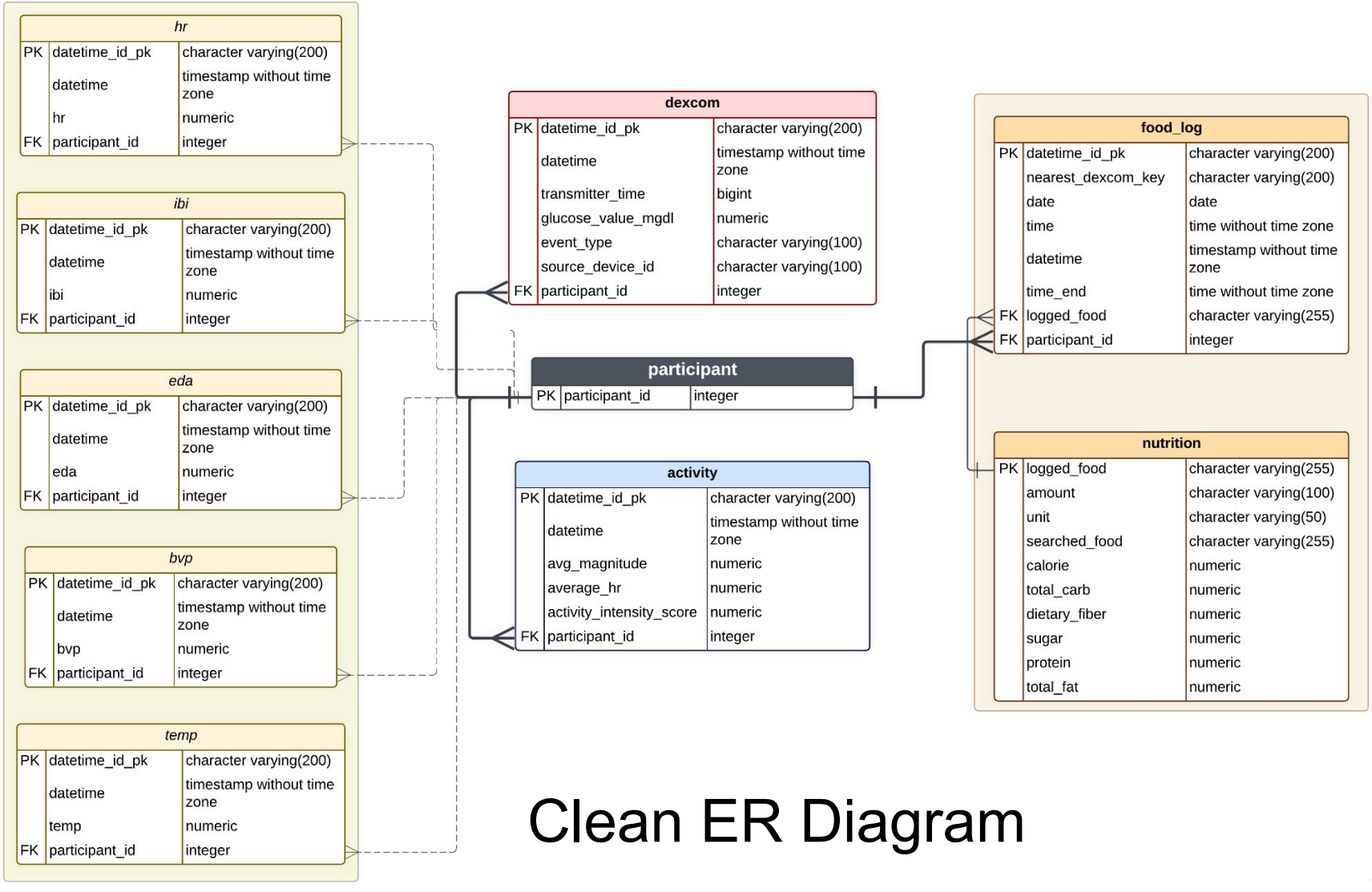
Perform analysis to answer research questions

Key activities

- Query and visualize glucose levels pre- and post-meals to track individual responses
- Analyze physical activity data to assess glucose stability
- Compare and interpret results across participants



Whole Diagram



Clean ER Diagram