

Data visualization

Bushra Haque

2025-07-12

This plot visualizes the relationship between physical activity (steps taken in the last 30 minutes) and current glucose levels:

```
library(ggplot2)

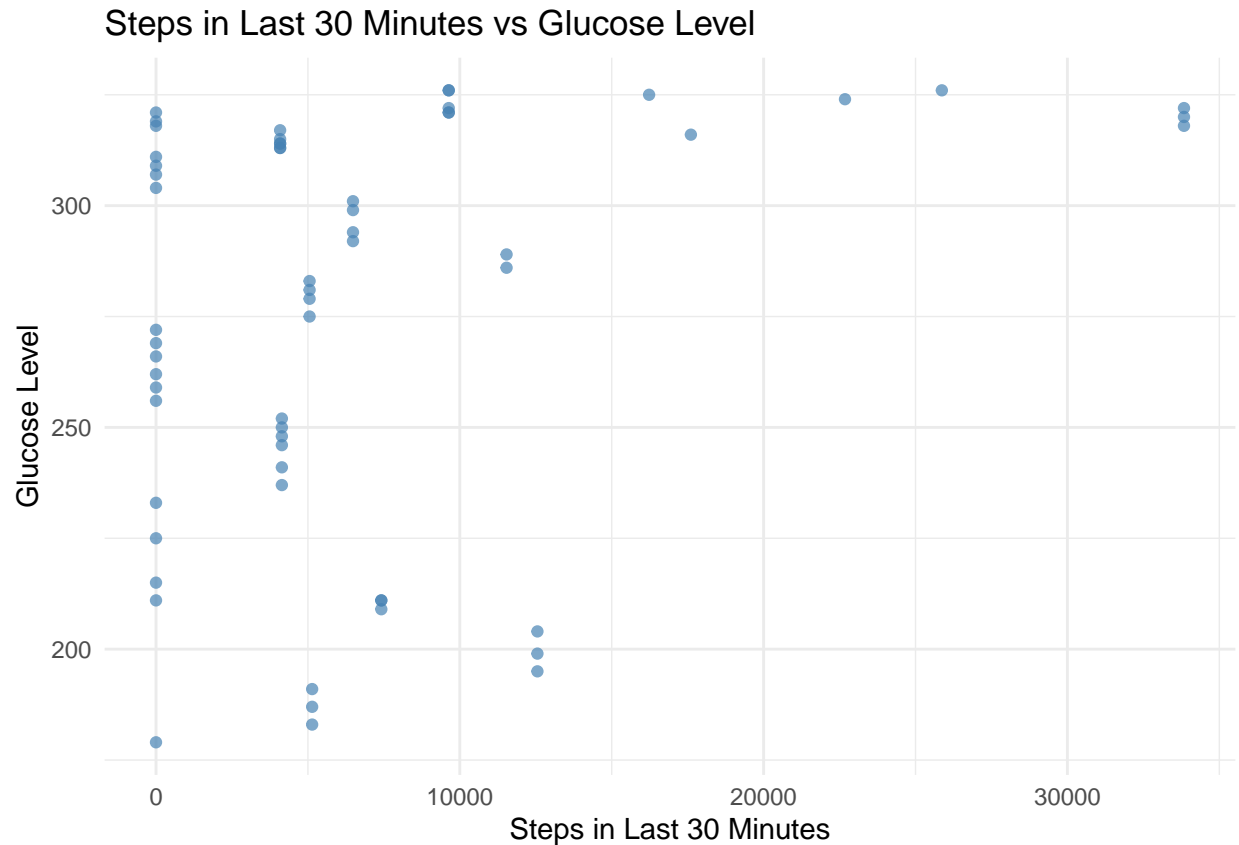
## Warning: package 'ggplot2' was built under R version 4.4.2

library(readr)

df <- read_csv("model_ready_dataset.csv")

## Rows: 61 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr  (2): date, time_of_day
## dbl  (2): level, steps_last_30min
## dtm   (1): timestamp
## time  (1): time
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

ggplot(df, aes(x = steps_last_30min, y = level)) +
  geom_point(color = "steelblue", alpha = 0.7) +
  labs(
    title = "Steps in Last 30 Minutes vs Glucose Level",
    x = "Steps in Last 30 Minutes",
    y = "Glucose Level"
  ) +
  theme_minimal()
```



As you can see the glucose level are ranging from different step levels. As 30 minutes is not enough time for a significant change, you can still see how it makes a gradual change in the plot. We also keep in mind that the absence of a strong short-term correlation may reflect the complexity of individual physiology and time lags between activity and measurable glucose changes, not a lack of benefit.

Box plot representing improvement:

```
library(ggplot2)
library(readr)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```

df <- read_csv("model_ready_dataset.csv")

## Rows: 61 Columns: 6

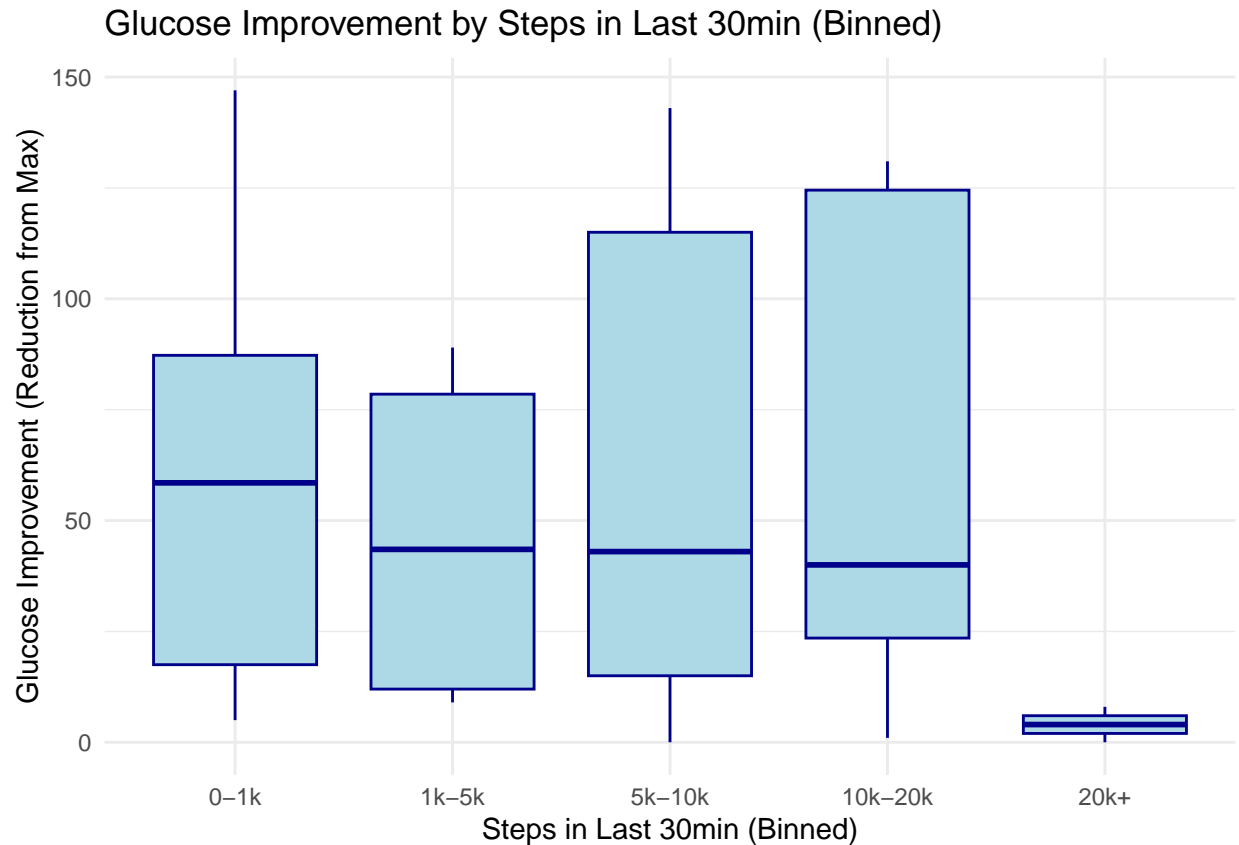
## -- Column specification -----
## Delimiter: ","
## chr  (2): date, time_of_day
## dbl  (2): level, steps_last_30min
## dtm  (1): timestamp
## time (1): time
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

df <- df %>%
  mutate(glucose_improvement = max(level, na.rm = TRUE) - level)

df <- df %>%
  mutate(steps_bin = cut(steps_last_30min,
                        breaks = c(-Inf, 1000, 5000, 10000, 20000, Inf),
                        labels = c("0-1k", "1k-5k", "5k-10k", "10k-20k", "20k+")))

ggplot(df, aes(x = steps_bin, y = glucose_improvement)) +
  geom_boxplot(fill = "lightblue", color = "darkblue") +
  labs(
    title = "Glucose Improvement by Steps in Last 30min (Binned)",
    x = "Steps in Last 30min (Binned)",
    y = "Glucose Improvement (Reduction from Max)"
  ) +
  theme_minimal()

```



The boxplot summarizes how glucose improvement varies across different levels of walking activity. While there is variability within each group, it suggests that moderate walking (1k–10k steps) tends to be associated with higher reductions in glucose from baseline. This reinforces the idea that physical activity is associated with better glucose control on average, even if individual measurements vary.

Regression relation:

```
library(ggplot2)
library(readr)

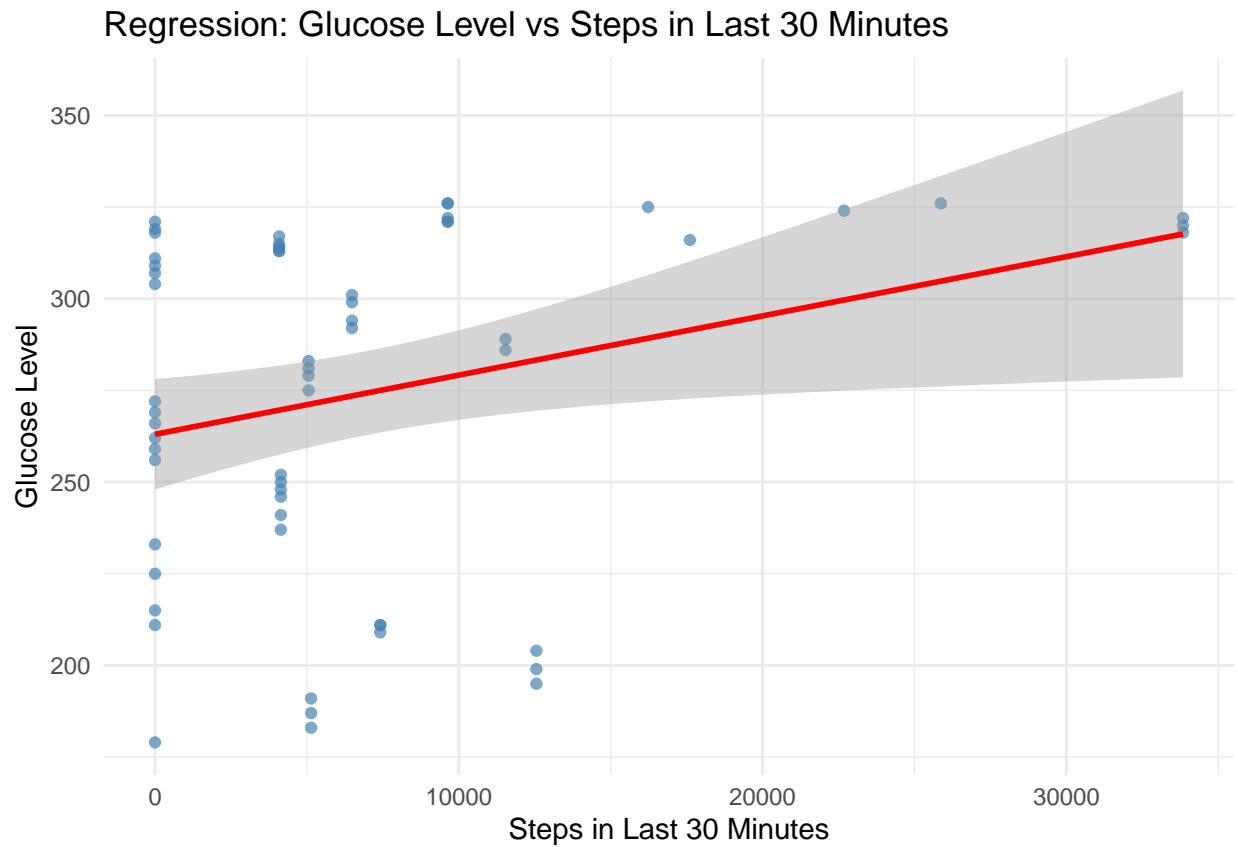
df <- read_csv("model_ready_dataset.csv")

## Rows: 61 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (2): date, time_of_day
## dbl (2): level, steps_last_30min
## dtm (1): timestamp
## time (1): time
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

ggplot(df, aes(x = steps_last_30min, y = level)) +
  geom_point(color = "steelblue", alpha = 0.7) +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
```

```
labs(
  title = "Regression: Glucose Level vs Steps in Last 30 Minutes",
  x = "Steps in Last 30 Minutes",
  y = "Glucose Level"
) +
theme_minimal()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



This regression suggests other complexity of glucose regulation: factors like food intake, insulin administration, and physiological lags between exercise and glucose response can obscure this short-term relationship.