Quantitative Activity Levels and Gestational Age at Delivery

Sara Kim

```
fitbit = readxl::read_xlsx("./data/final.xlsx") %>%
  janitor::clean_names() %>%
  filter(ga_delivery > 200 | day > ga_delivery) %>%
  #removed one case of miscarriage
  mutate(id = as.factor(id))
```

Goal: The primary exposure and outcome will be the mean number of steps/day and the risk of spontaneous preterm delivery (<37 weeks), respectively. We will assess if the mean number of steps/day will differ between women that spontaneously deliver preterm versus those that deliver at term.

EDA

```
fitbit[,c(3:5)] %>%
 summary()
```

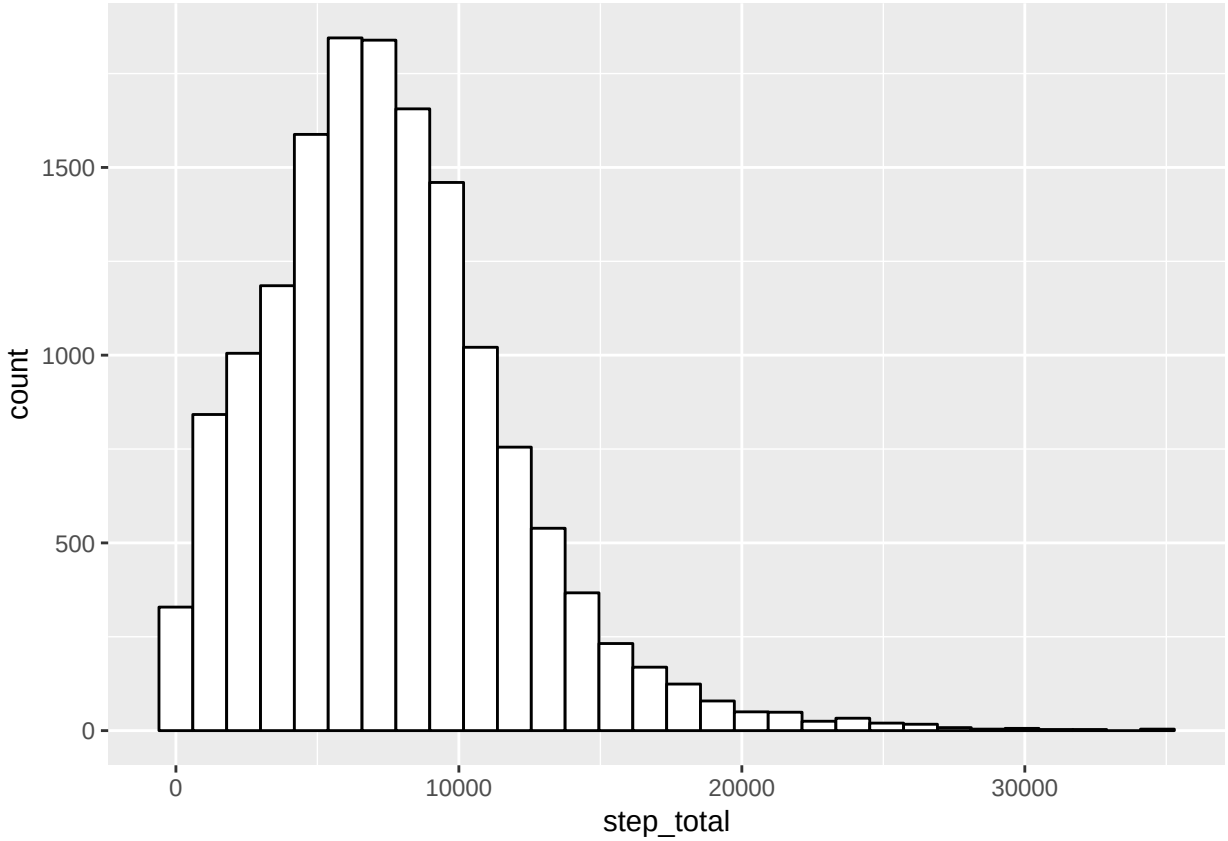
```
##       day          step_total     ga_delivery
##  Min.   : 43.0   Min.   :  101   Min.   :225.0
##  1st Qu.:136.0   1st Qu.: 4546   1st Qu.:269.0
##  Median :177.0   Median : 7104   Median :275.0
##  Mean   :176.9   Mean   : 7547   Mean   :274.4
##  3rd Qu.:220.0   3rd Qu.: 9920   3rd Qu.:282.0
##  Max.   :288.0   Max.   :34788   Max.   :289.0
```

Comparison: United States: 5,117 steps. This is about 2.5 miles or about 4 kilometers each day. Switzerland: 9,650 steps. This is about 4.8 miles or 8 kilometers each day. Japan: 7,168 steps. This is about 3.5 miles or 6 kilometers each day.

We can notice large variance in steps and higher steps than us mean.

```
ggplot(fitbit, aes(x=step_total)) +
  geom_histogram(color ="black", fill="white")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# removing step_total with less than 300\
fitbit = fitbit %>%
  filter(step_total > 300)

fitbit %>%
  group_by(term) %>%
  summarize(median_steps = median(step_total),
            mean_steps = mean(step_total),
            sd_steps = sd(step_total)) %>%
  knitr::kable()
```

term	median_steps	mean_steps	sd_steps
PRE	7564	7677.422	3985.331
TERM	7127	7613.137	4348.893
Non-pre	term has a larg	er variance b	ut both groups have very similar number of steps.

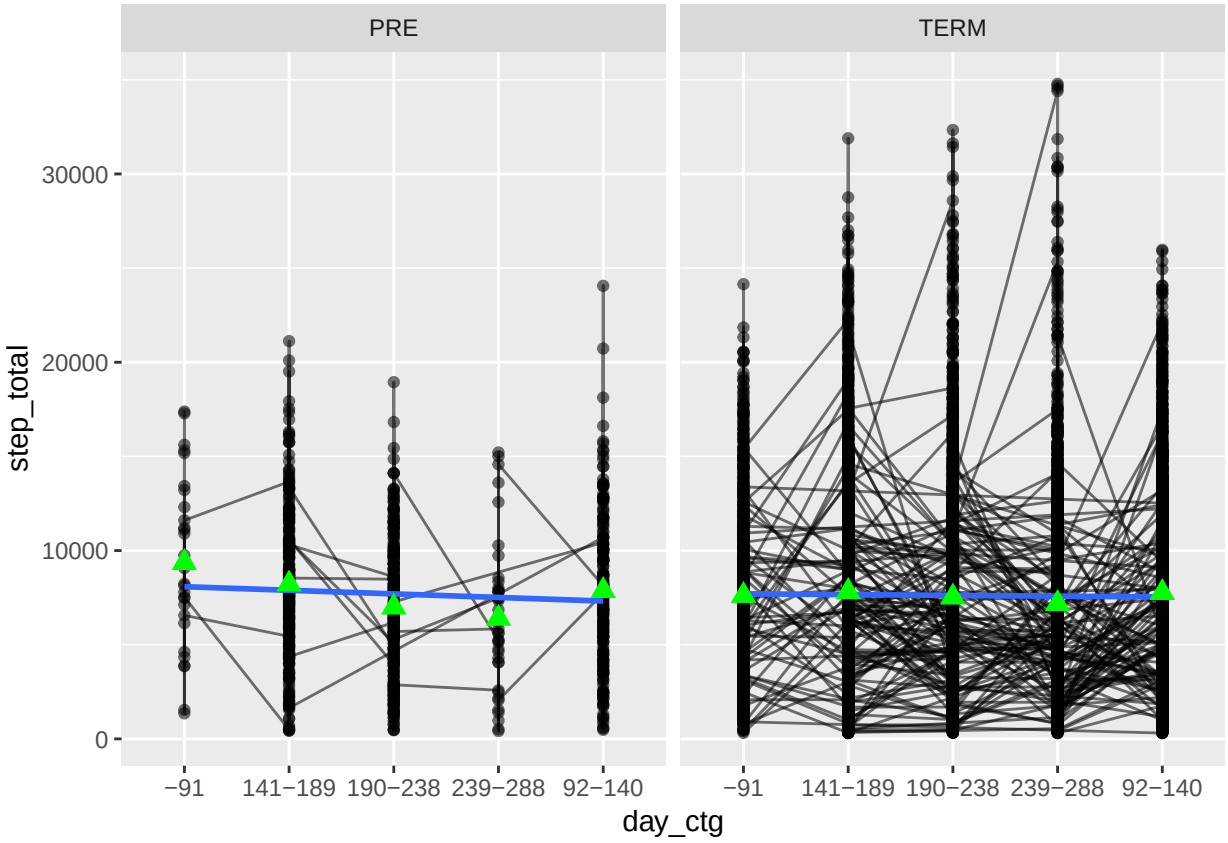
Longitudinal Effect

I sectioned the days into 5 windows.

```
fitbit = fitbit %>%
  mutate(
      day_ctg = case_when(
      day < 92 ~ "-91",
      day %in% 92:140 ~ "92-140",
```

```
      day %in% 141:189 ~ "141-189",
      day %in% 190:238 ~ "190-238",
      day %in% 239:288 ~ "239-288"
    ),
    day_ctg = fct_relevel(day_ctg, "-91"))
```

```
fitbit %>%
  ggplot(aes(x = day_ctg, y = step_total, group = id)) + geom_point(aes(alpha = 0.2)) + geom_line(aes(al
    stat_summary(aes(group = 1), geom = "point", color = "green", fun.y = mean, shape = 17, size = 3) +
    facet_grid(. ~ term)  + theme(legend.position = "none")
```



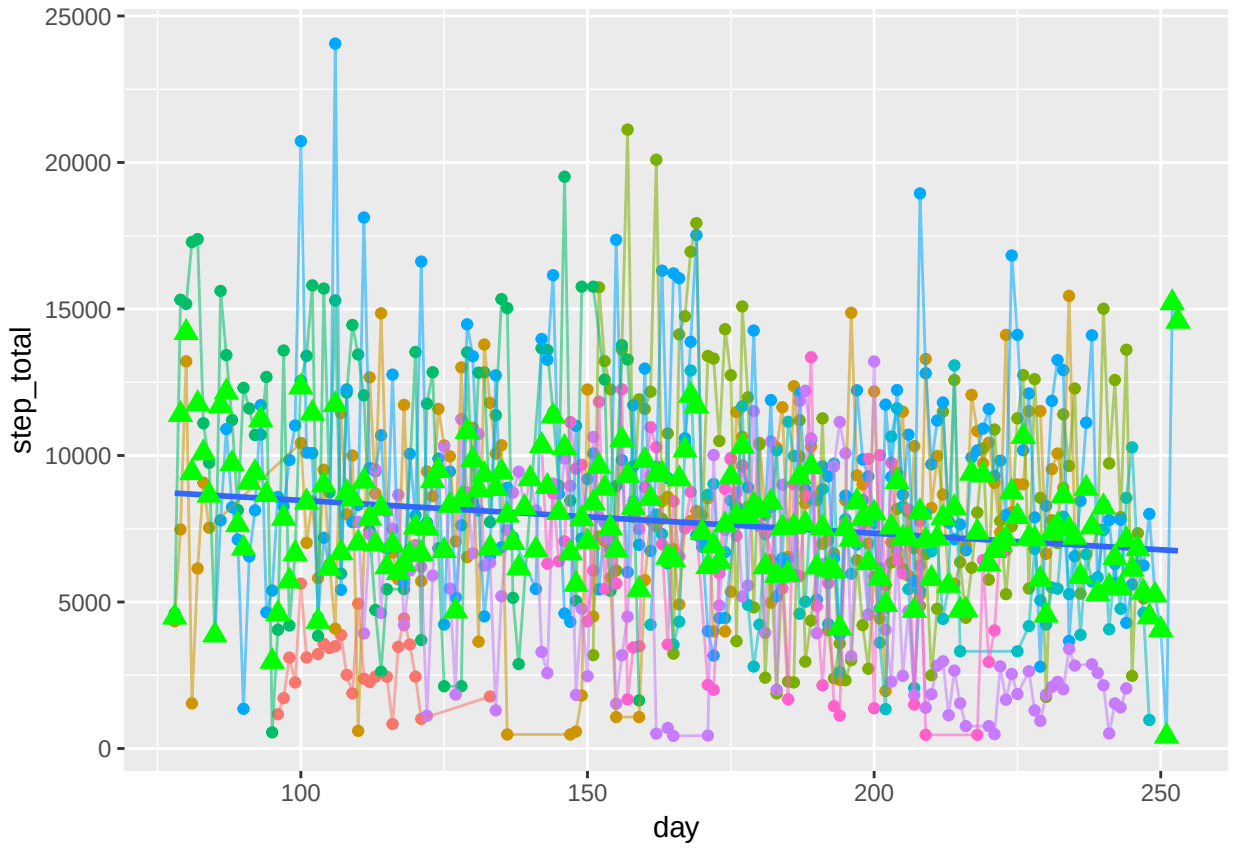
I windowed the gestational age (day) into 5 windows.

The vertical points we see above are the steps data observed in these five time windows (day < 92, $91 <$ day $< 141, \ldots, 238 <$ day < 289).

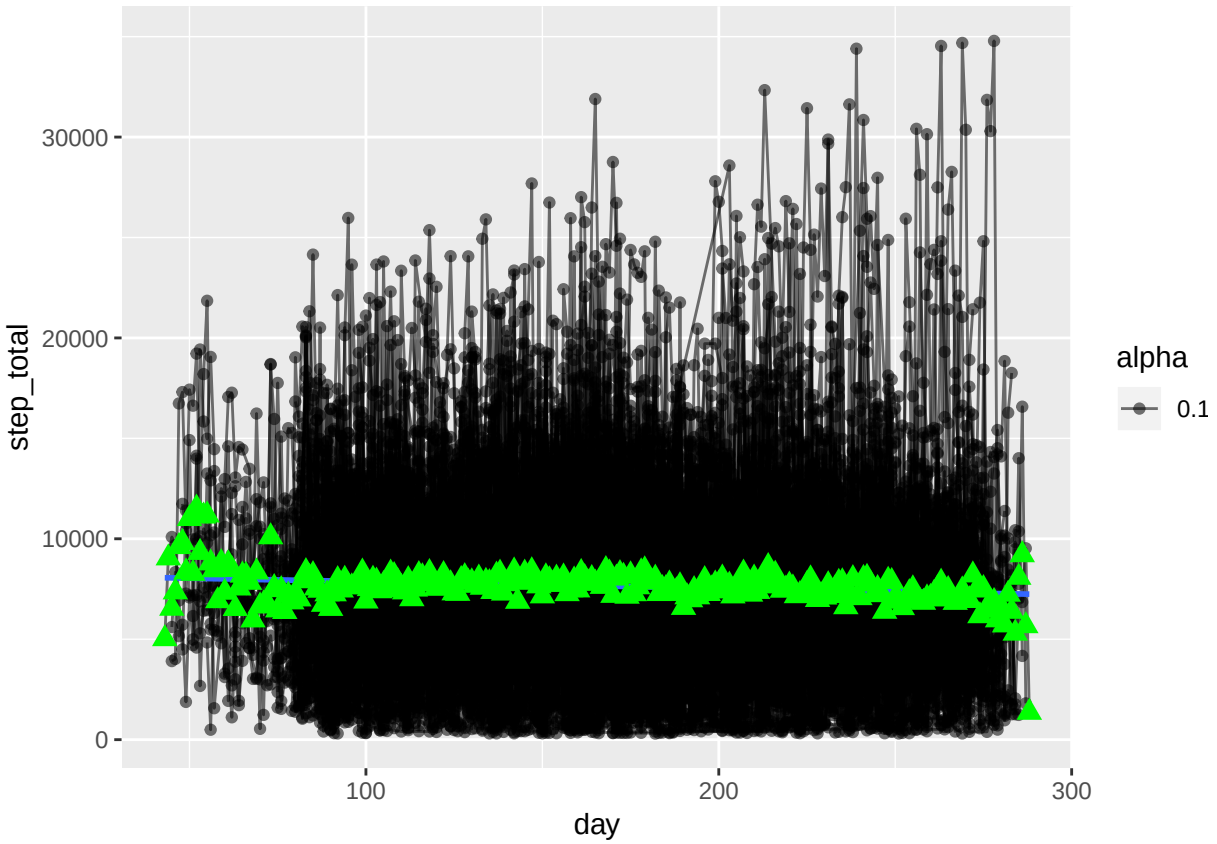
It does not seem like there is a noticeable difference in the steps based on the gestational age.

I will also plotted the days vs step_total without grouping the days.

```
fitbit %>%
  filter(term == "PRE") %>%
  ggplot(aes(x = day, y = step_total, group = id, color = id)) + geom_point() + geom_line(aes(alpha = 0
    stat_summary(aes(group = 1), geom = "point", color = "green", fun.y = mean, shape = 17, size = 3)
```




```
fitbit %>%
  filter(term == "TERM") %>%
  ggplot(aes(x = day, y = step_total, group = id)) + geom_point(aes(alpha = 0.1)) + geom_line(aes(alpha
    stat_summary(aes(group = 1), geom = "point", color = "green", fun.y = mean, shape = 17, size = 3)
```



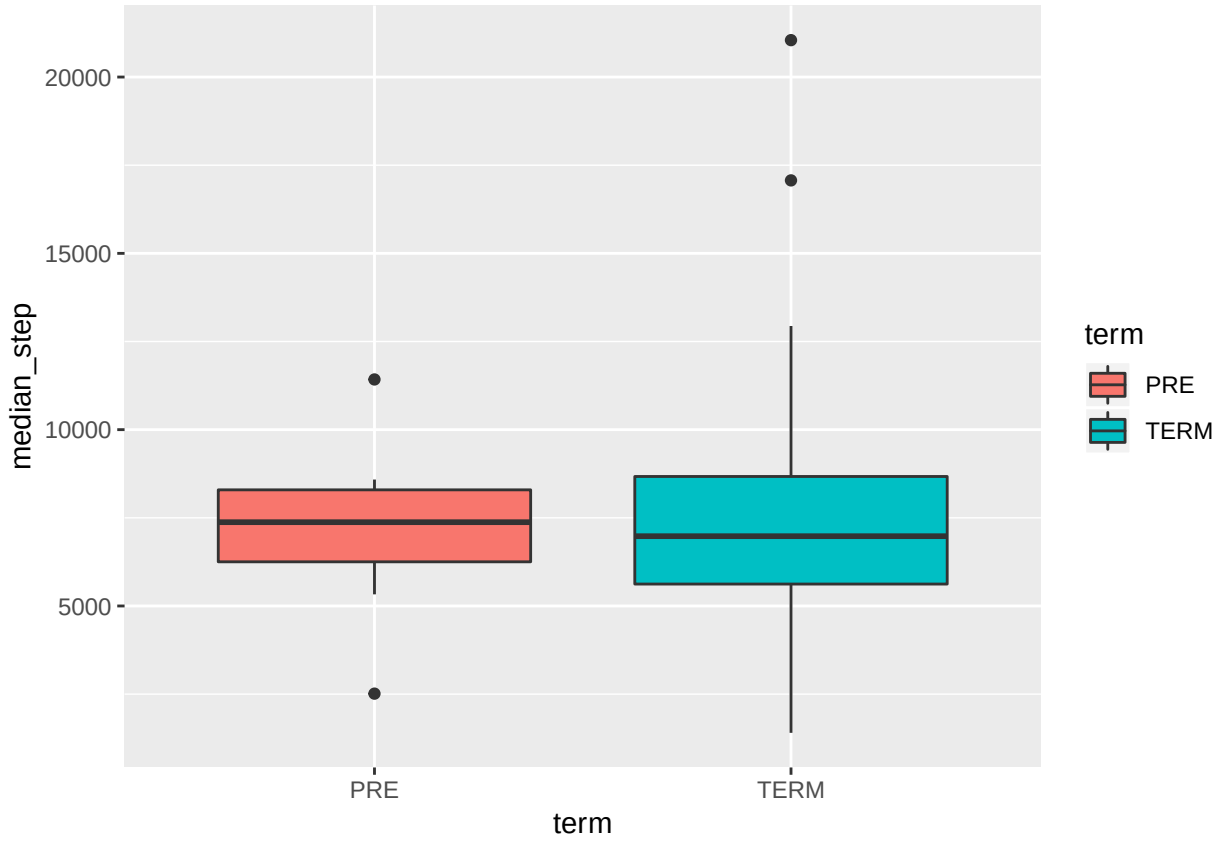
We do see a slight downward trend in the steps for those in the pre-term birth group. However, it seems like indivisuals' observations are not spread across the time evenly. For instance, the purple line we see above is not observerd before the 100 gestational day. That patient's observed values have extremely low values and it seems to be causing to lower the regression line in the later time points.

Median Steps per Day

```
fitbit_median = fitbit %>%
  group_by(id, term) %>%
  summarise(median_step = median(step_total),
            mean_step = mean(step_total),
            ga_delivery = mean(ga_delivery))

fitbit_median %>%
  group_by(term) %>%
  summarise(n = n(),
            percent = n()/nrow(fitbit_median)) %>%
  knitr::kable()
```

term	n	percent
PRE	8	0.0620155
TERM	121	0.9379845

```
ggplot(fitbit_median, aes(term,median_step, fill=term)) +
    geom_boxplot()
```

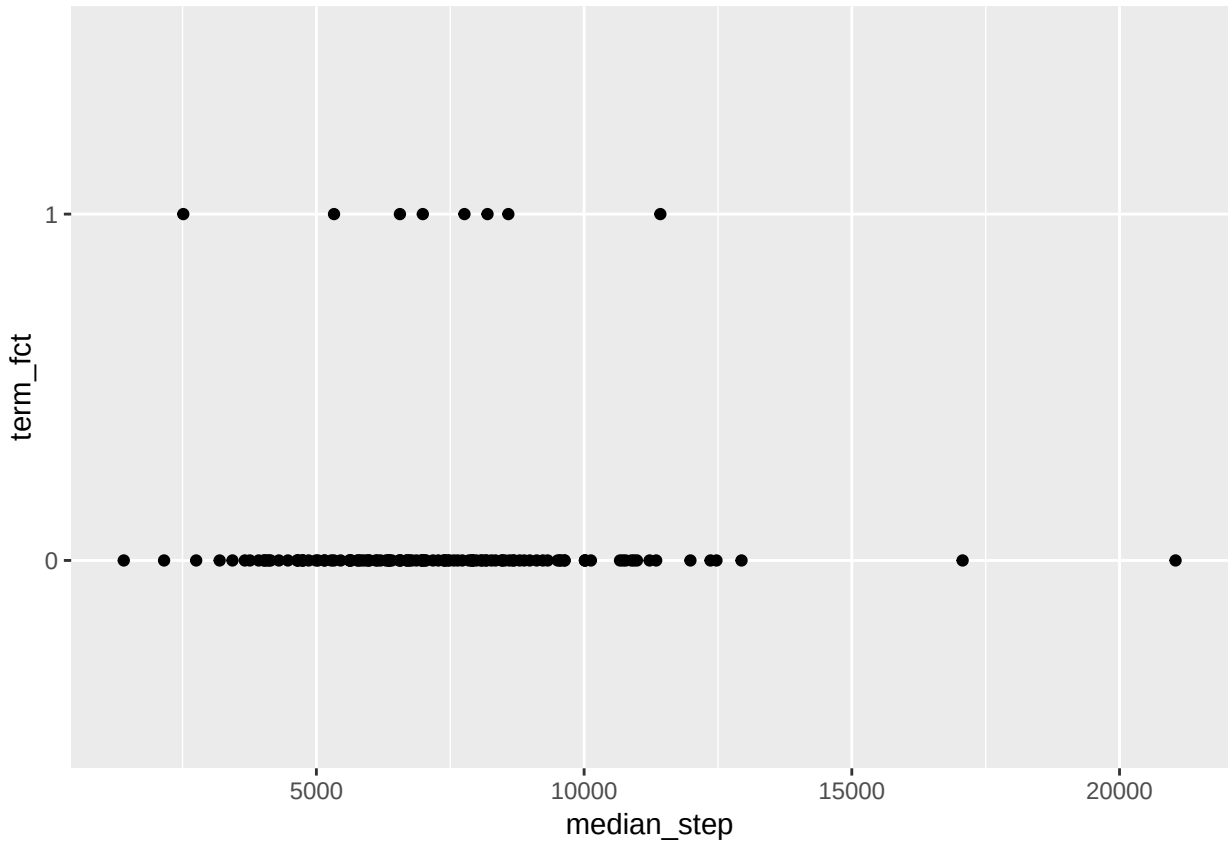



```
fitbit_median %>%
  group_by(term) %>%
  summarise(n = n(),
            median_step= median(median_step))
```

```
## # A tibble: 2 x 3
##   term      n median_step
##   <chr> <int>       <dbl>
## 1 PRE       8       7376.
## 2 TERM    121       6979
```

Looking at the median steps for each indivisual, the median of the median steps seems similar between the two groups. The median value of the pre-term birth cases is actually slightly higher (794 steps).

```
term_fct = as.factor(ifelse(fitbit_median$term == "TERM", 0, 1))
fitbit_median["term_fct"] = term_fct

fitbit_median %>%
  ggplot(aes(x=median_step, y=term_fct)) +
  geom_point()
```



logistic regression

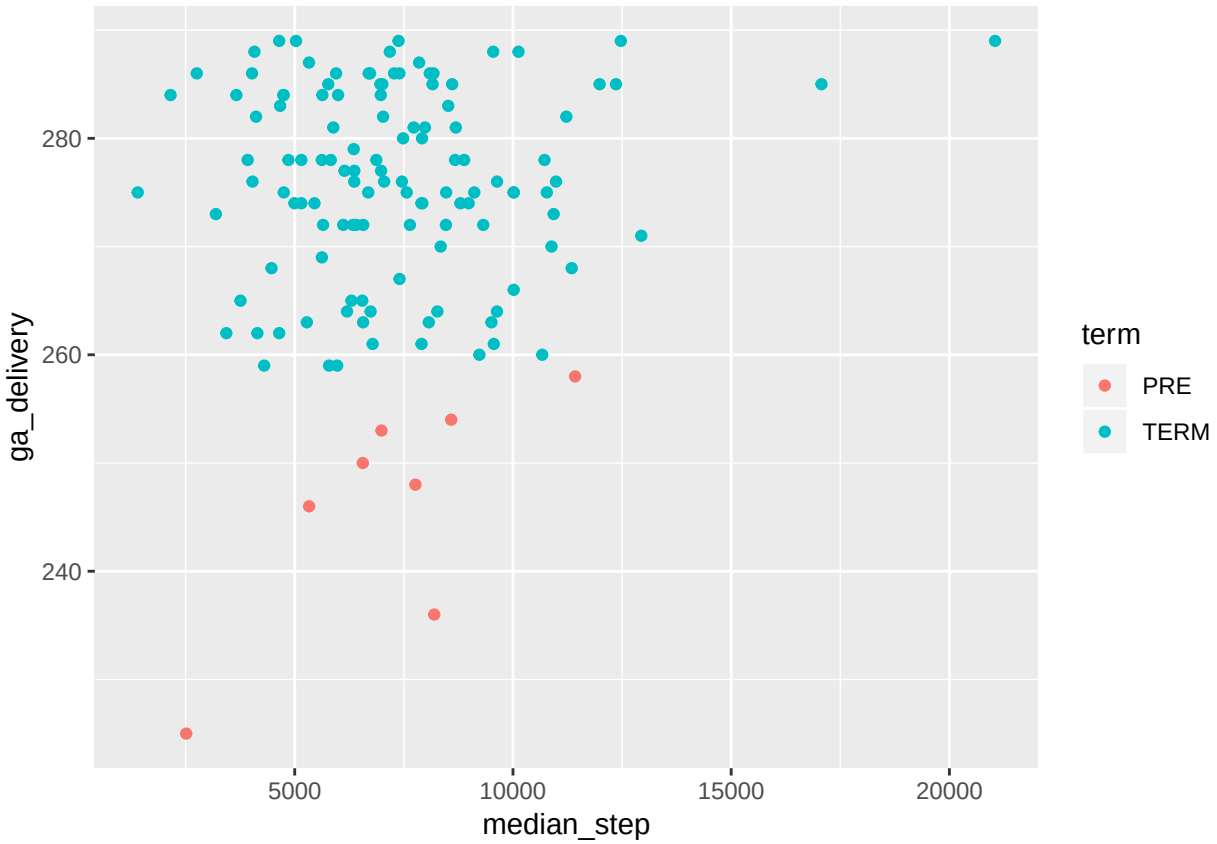
```
glm <- glm(term_fct ~ median_step, family=binomial(link='logit'), data = fitbit_median)

glm  %>%
  broom::tidy() %>%
  knitr::kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-2.568	1.05	-2.447	0.014
median_step	0.000	0.00	-0.149	0.882

there is no statistically significant evidence that there is a linear relationship between log odds of pre-term birth and median steps per day.

Relationship Between Delivery Date and Median Steps Per Day

```
fitbit_median %>%
  ggplot(aes(x=median_step, y= ga_delivery, color = term)) +
  geom_point()
```




```
res <- cor.test(fitbit_median$median_step, fitbit_median$ga_delivery,
                    method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  fitbit_median$median_step and fitbit_median$ga_delivery
## t = 1.0884, df = 127, p-value = 0.2785
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.07801452  0.26459375
## sample estimates:
##       cor
## 0.09613627
```

```
res$p.value
```

```
## [1] 0.2784611
```

At significance level of 5%, we do not reject the null hypothesis and can conclude that there is not significant correlation between delivery date and median steps per day.

Fisher's Exact Test

```
# calculating cutoff value
fitbit_median$median_step %>% median()
```

```
## [1] 6987.5
```

```r
# 6987.5

fitbit_median = fitbit_median %>%
  mutate(step_group = ifelse(median_step < 6987.5, "low", "high"))

step_mat = fitbit_median %>%
  group_by(step_group, term) %>%
  summarise(n = n()) %>%
  pull(n) %>%
  matrix(nrow = 2)

fisher.test(step_mat)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  step_mat
## p-value = 0.7178
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   0.3124644 11.3509590
## sample estimates:
## odds ratio
##   1.687636
```

The p-value calculated for the test does not provide any evidence against the assumption of independence. In this example this means that we cannot confidently claim any difference in median_steps for the pre-term birth group and term-birth group.