Fitbit Analysis

Sara Kim

```
fitbit = readxl::read_xlsx("./data/final.xlsx") %>%
  janitor::clean_names()
```

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.   :191.0
##  1st Qu.:136.0   1st Qu.: 4554   1st Qu.:268.0
##  Median :176.0   Median : 7120   Median :275.0
##  Mean   :176.7   Mean   : 7559   Mean   :273.9
##  3rd Qu.:220.0   3rd Qu.: 9963   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`.
```
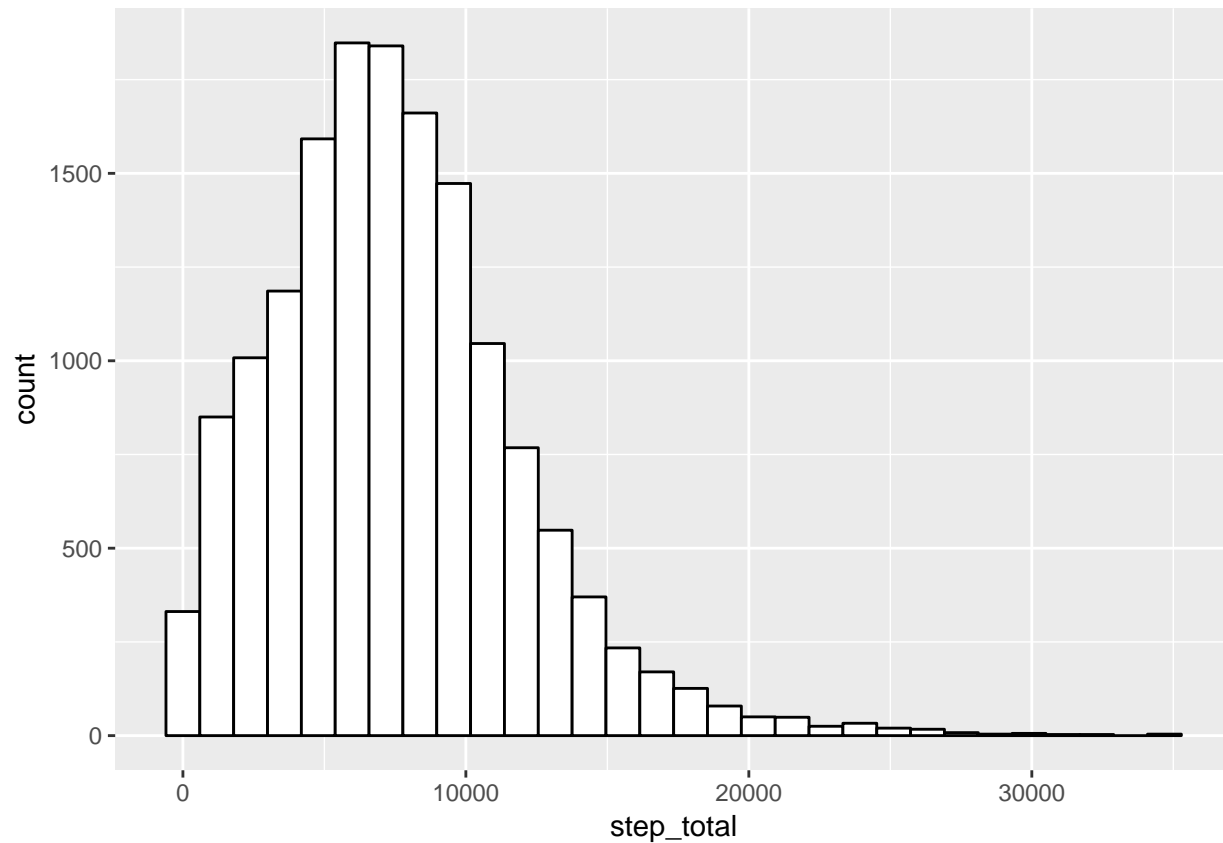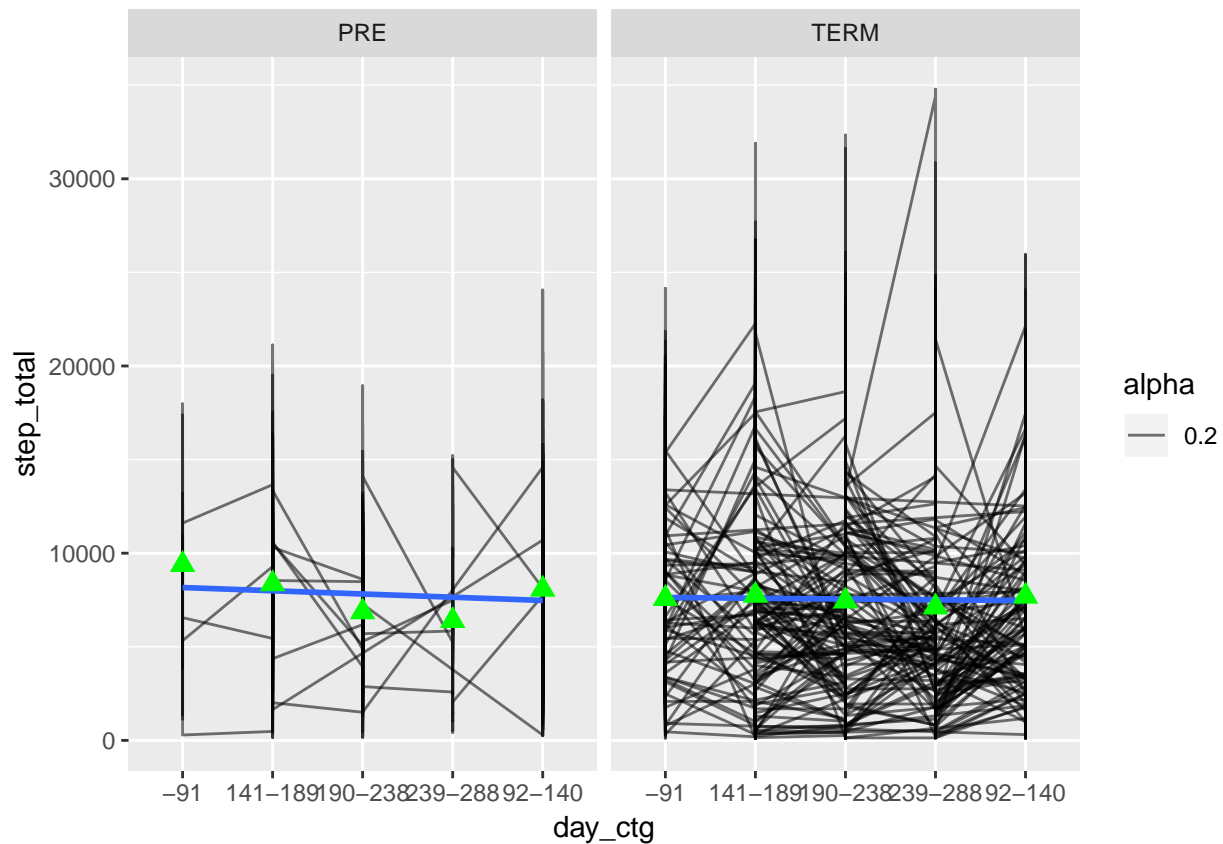
```
fitbit %>%
  group_by(term) %>%
  summarize(median_steps = median(step_total),
        mean_steps = mean(step_total),
        sd_steps = sd(step_total)) %>%
  knitr::kable()
```

term	median_steps	mean_steps	sd_steps
PRE	7768	7802.261	4092.346
TERM	7090	7544.561	4386.590
Non-pre	term has a larg	er variance b	ut both groups have very similar number of steps.

Longitudinal Effect

```
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_line(aes(alpha = 0.2)) + stat_smooth(aes(g
    stat_summary(aes(group = 1), geom = "point", color = "green", fun.y = mean, shape = 17, size = 3) +
    facet_grid(. ~ term)
```


I windowed the gestational age (day) into 5 windows. It does not seem like there is a difference in the steps based on the gestational age.
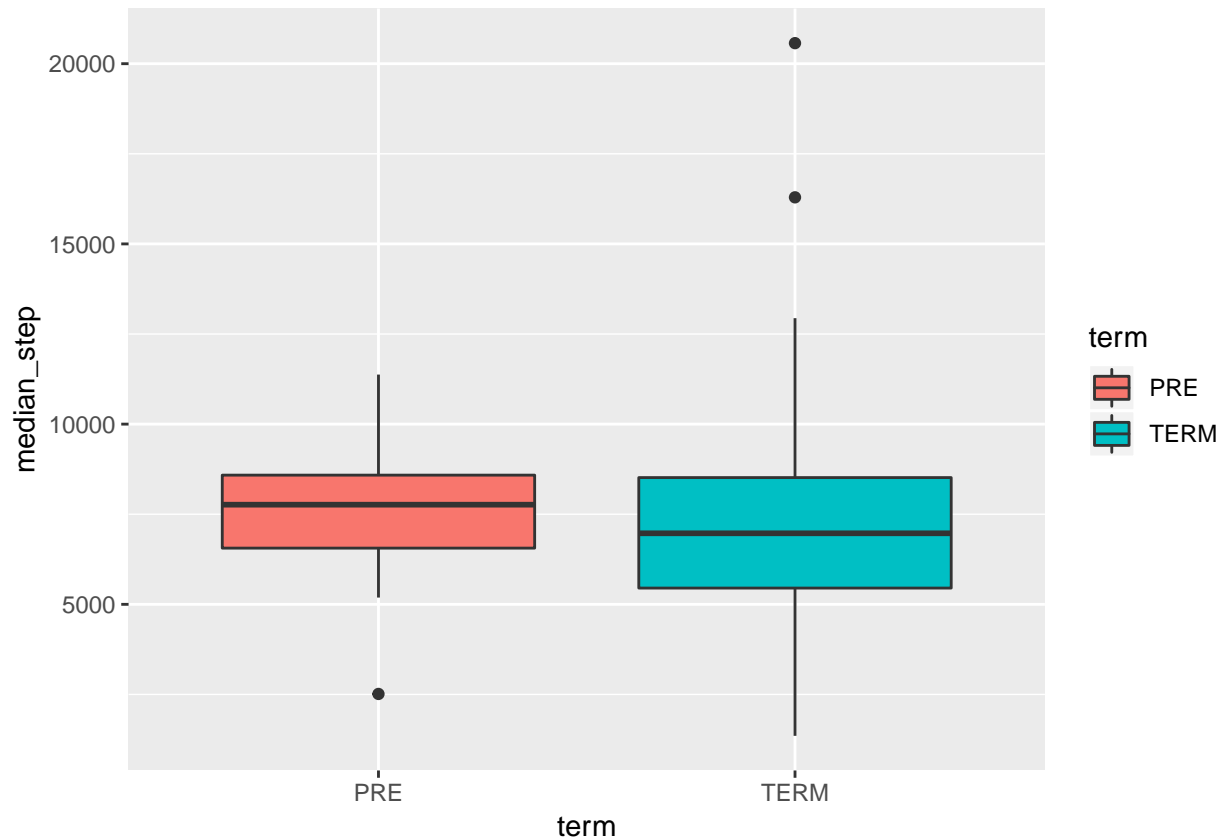
Median Steps per Day

```
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	9	0.0692308
TERM	121	0.9307692

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



```r
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       9        7765
## 2 TERM    121        6971
```

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).

```r
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()
```
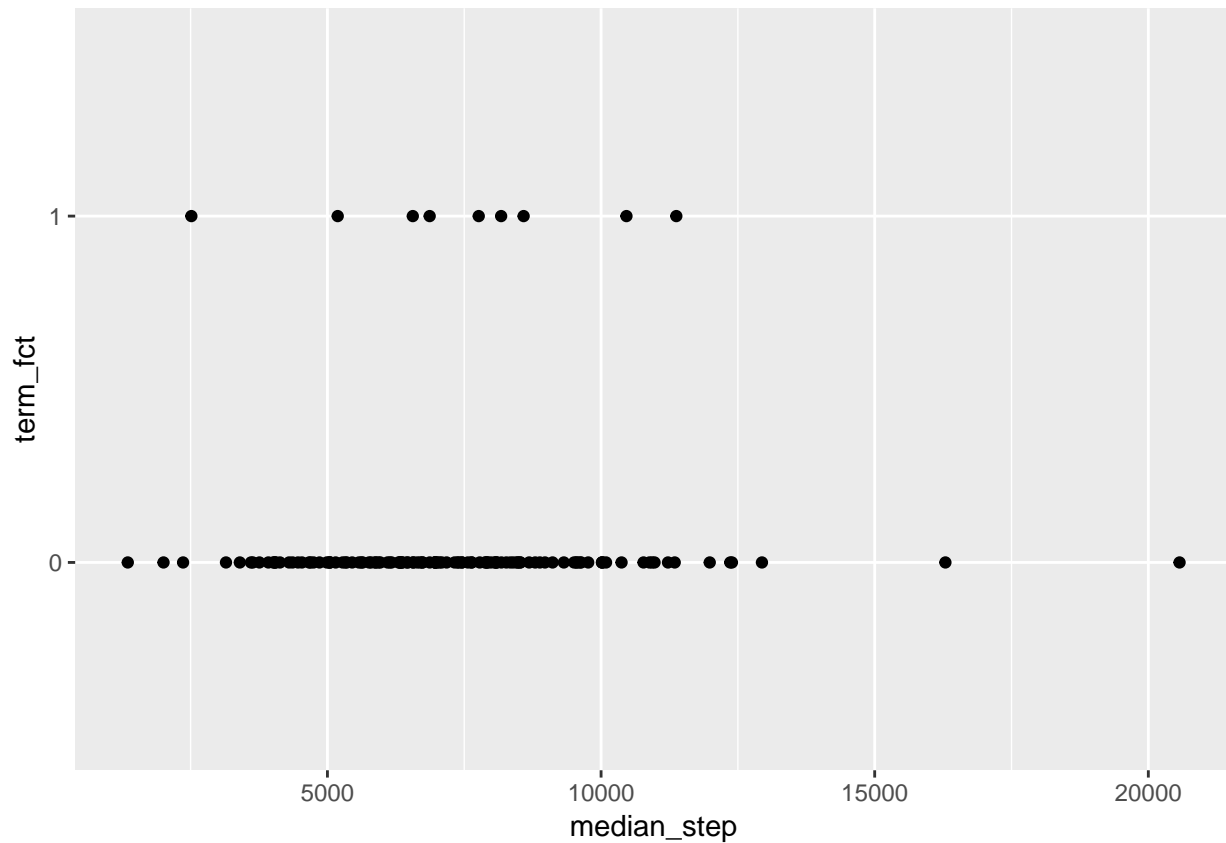
logistic regression

```
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)	-2.841	0.97	-2.928	0.003
median_step	0.000	0.00	0.271	0.786

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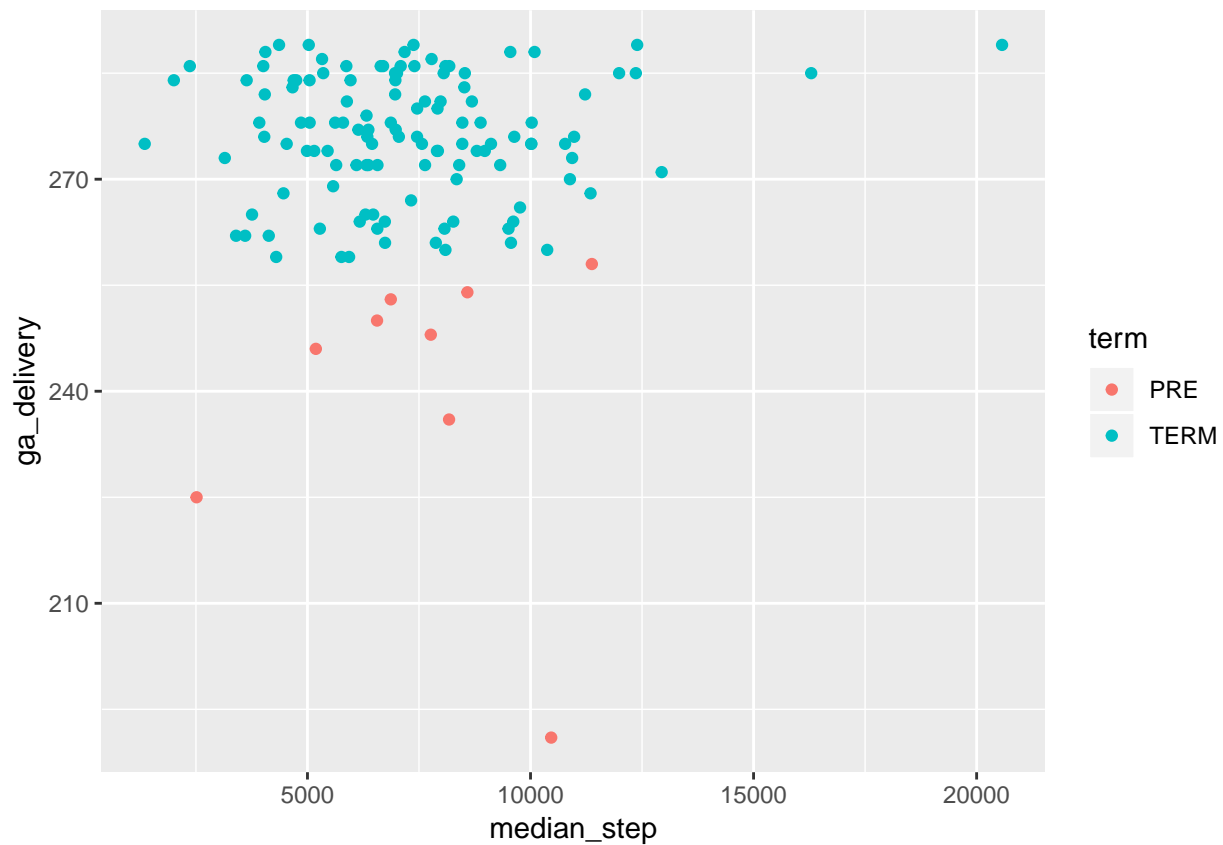
```
res <- cor.test(fitbit_median$median_step, fitbit_median$ga_delivery,
                method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  fitbit_median$median_step and fitbit_median$ga_delivery
## t = 0.27281, df = 128, p-value = 0.7854
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1486968  0.1954812
## sample estimates:
##        cor
## 0.02410654
```

```
res$p.value
```

```
## [1] 0.7854361
```