

# Visualizing Wearable Activity Tracker Data and Cardiometabolic Risk Factors with ggplot

---

Meike Niederhausen  
OHSU-PSU School of Public Health  
*niederha@ohsu.edu*

Research supported by the OHSU Knight Cardiovascular Institute

Cascadia R Conference, Portland, OR  
June 3, 2017

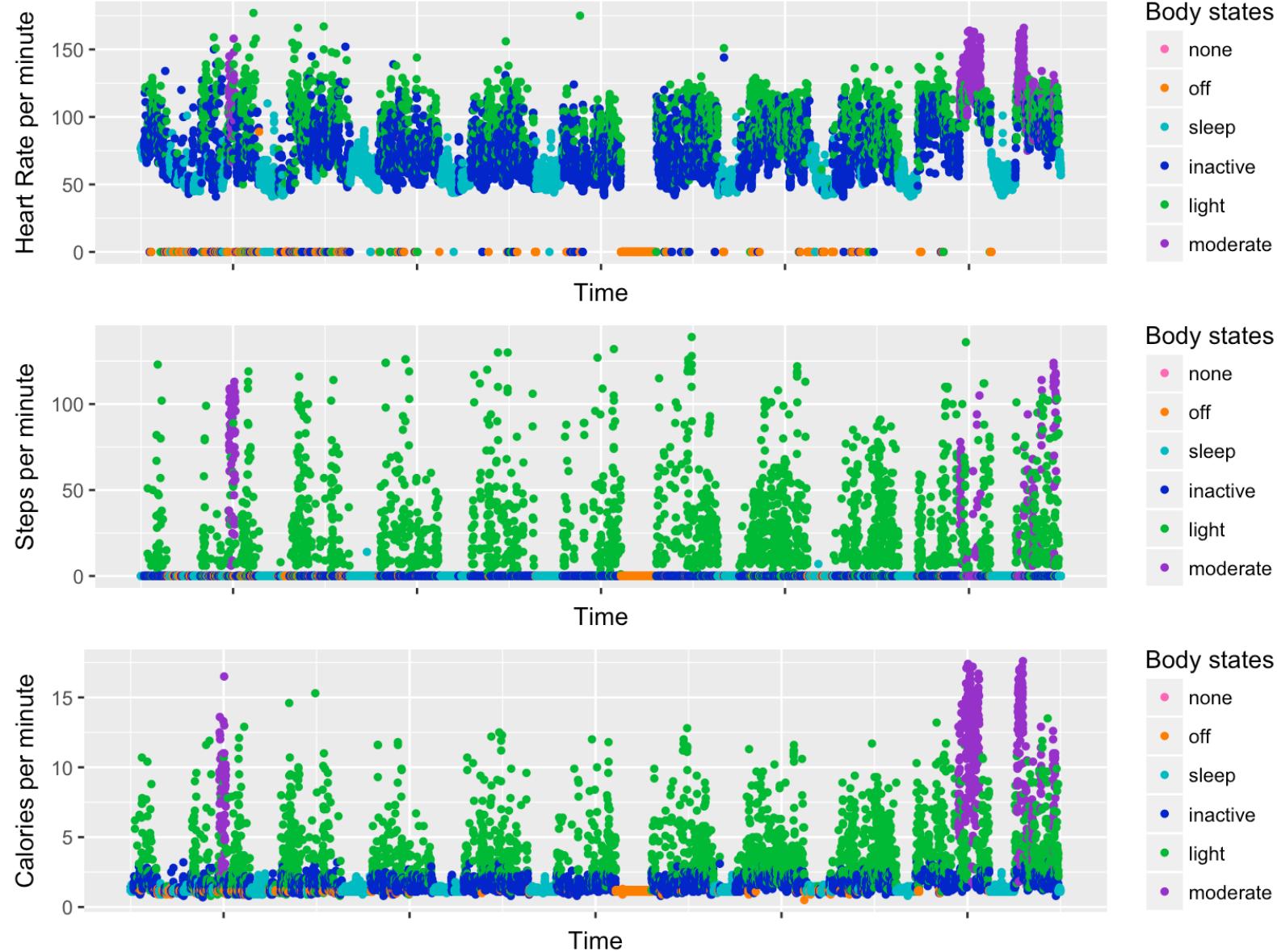
# Outline

- Introduction to the data
- Visualizing
  - the “raw” fitness tracker data
  - trends in steps per day
  - associations between different variables
  - correlations between activity and cardiometabolic risk factors
  - regression models comparing associations between activity and risk factors between ethnic groups
  - times of day and week with most frequent bouts of moderate activity

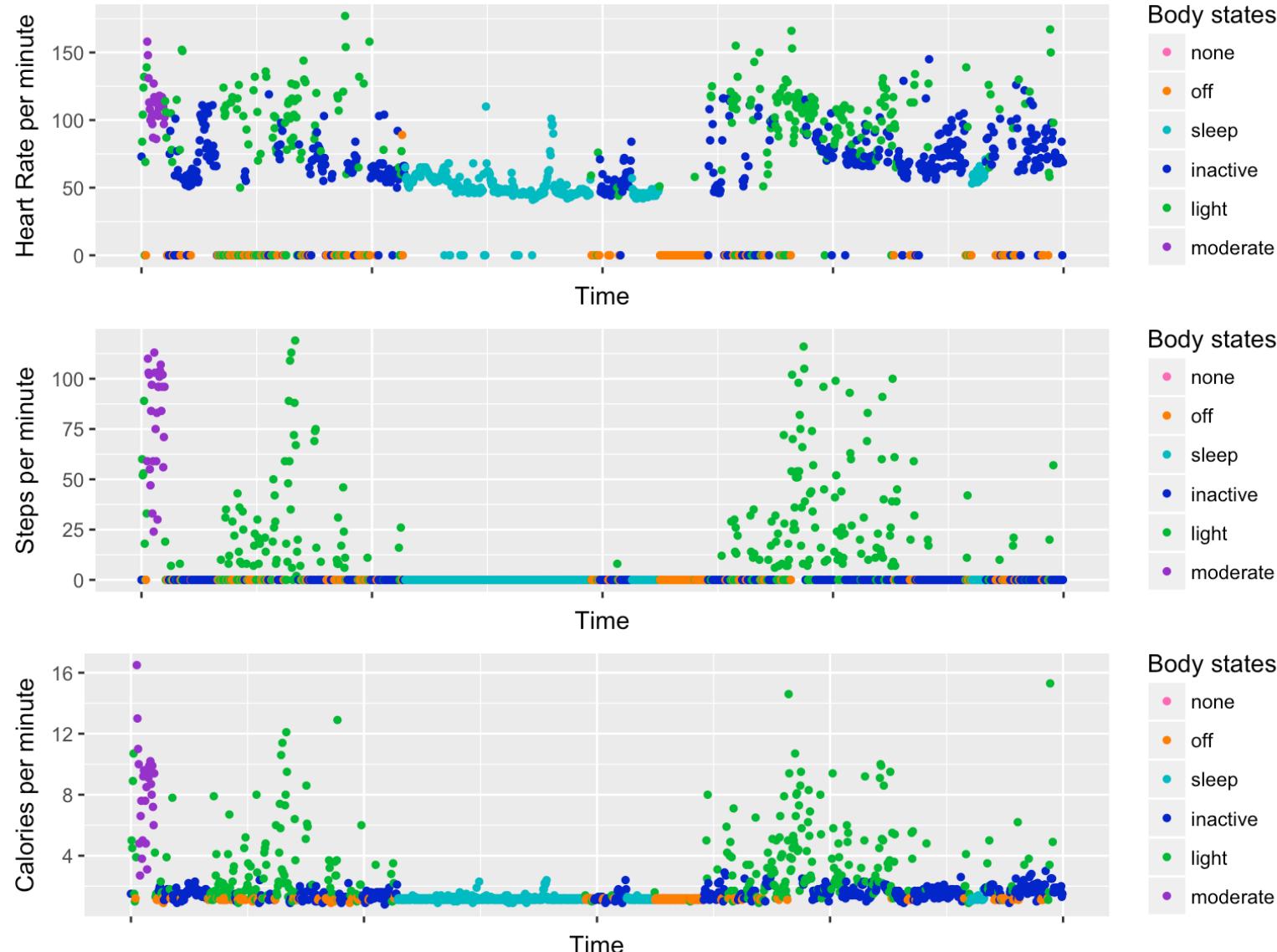
# The You 24x7 study

- 431 participants from same workplace
  - About 60% male
- 6-month study
  - Fitness tracker collected minute-by-minute activity data
    - Heart rate
    - Steps
    - Calories
    - Body states
    - Workout states
  - Baseline and end of study
    - Height, weight, blood pressure
    - Lab tests
      - Such as cholesterol levels, glucose, HbA1c
    - Online surveys
  - Electronic Health Records

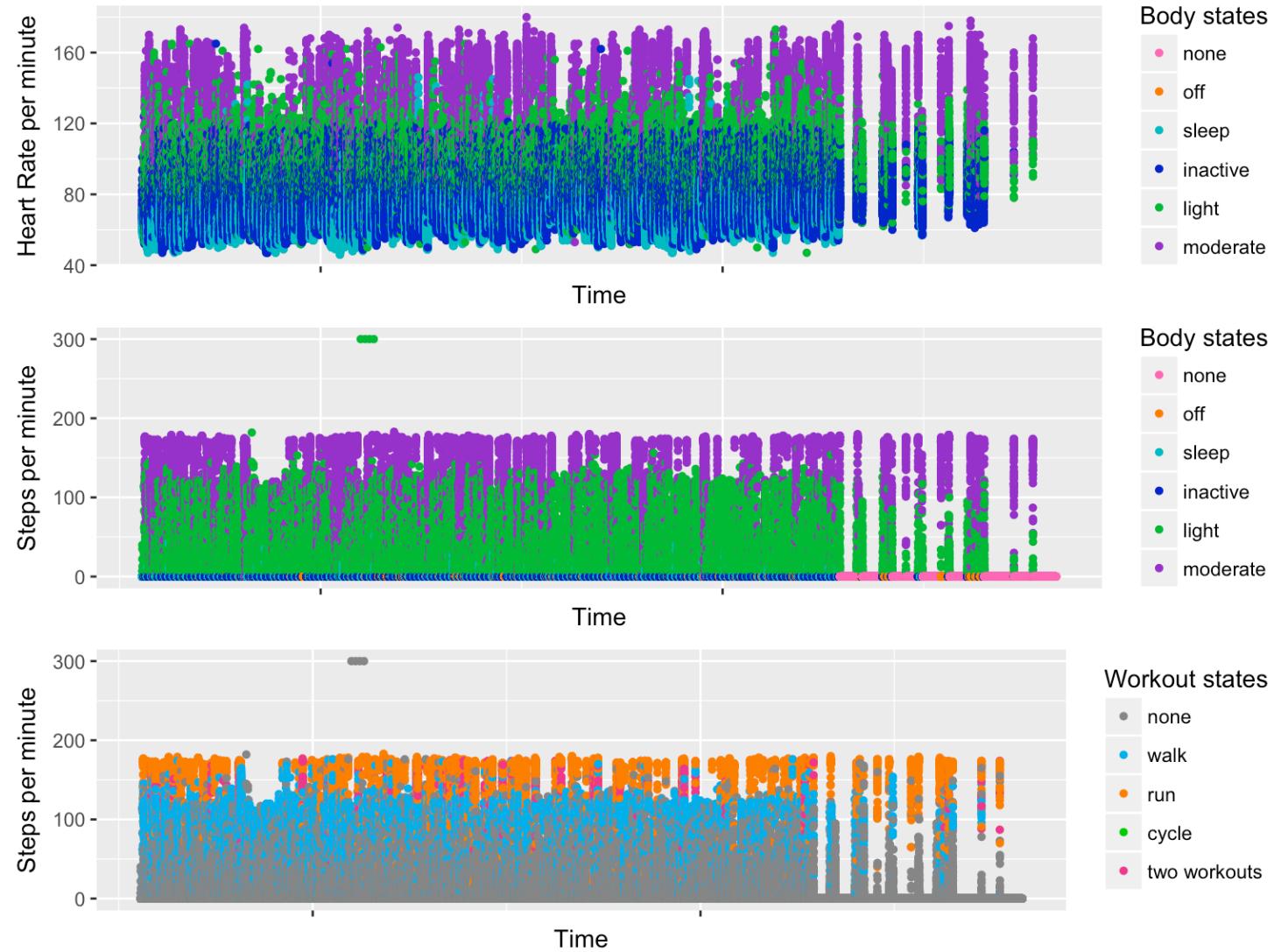
# Fitness tracker data (missing values = 0)



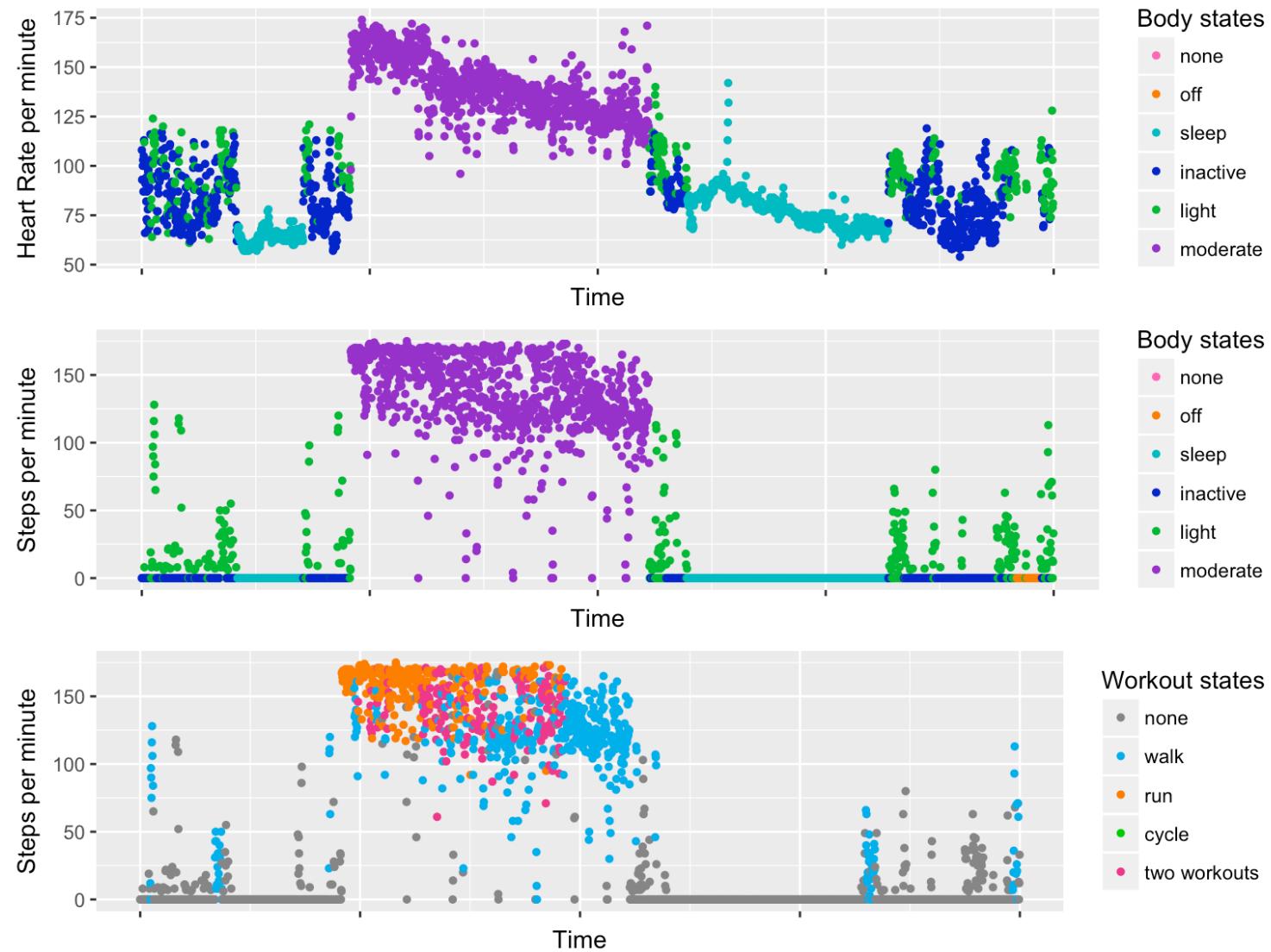
# Frequent off wrist periods in 24 hours: *jiggling*



# A highly active ID over study period



# Highly active ID zoomed in (2 days)



# Code: One ID, 10 days of minute-by-minute data

```
```{r}
cols.bd = c("hotpink", "darkorange1", "#00BFC4", "blue3", "#00BA38", "darkorchid3")

id_10days_HR_bodystates <-
  ggplot(id.week, aes(x=minute_time.utc, y=pulse, colour=bodystate)) +
  geom_point(size=1.1) + xlab("Time") + ylab("Heart Rate per minute") +
  scale_colour_manual(drop=TRUE, limits = levels(id.v2$bodystate), values= cols.bd, name="Body states") +
  theme(axis.text.x = element_blank())

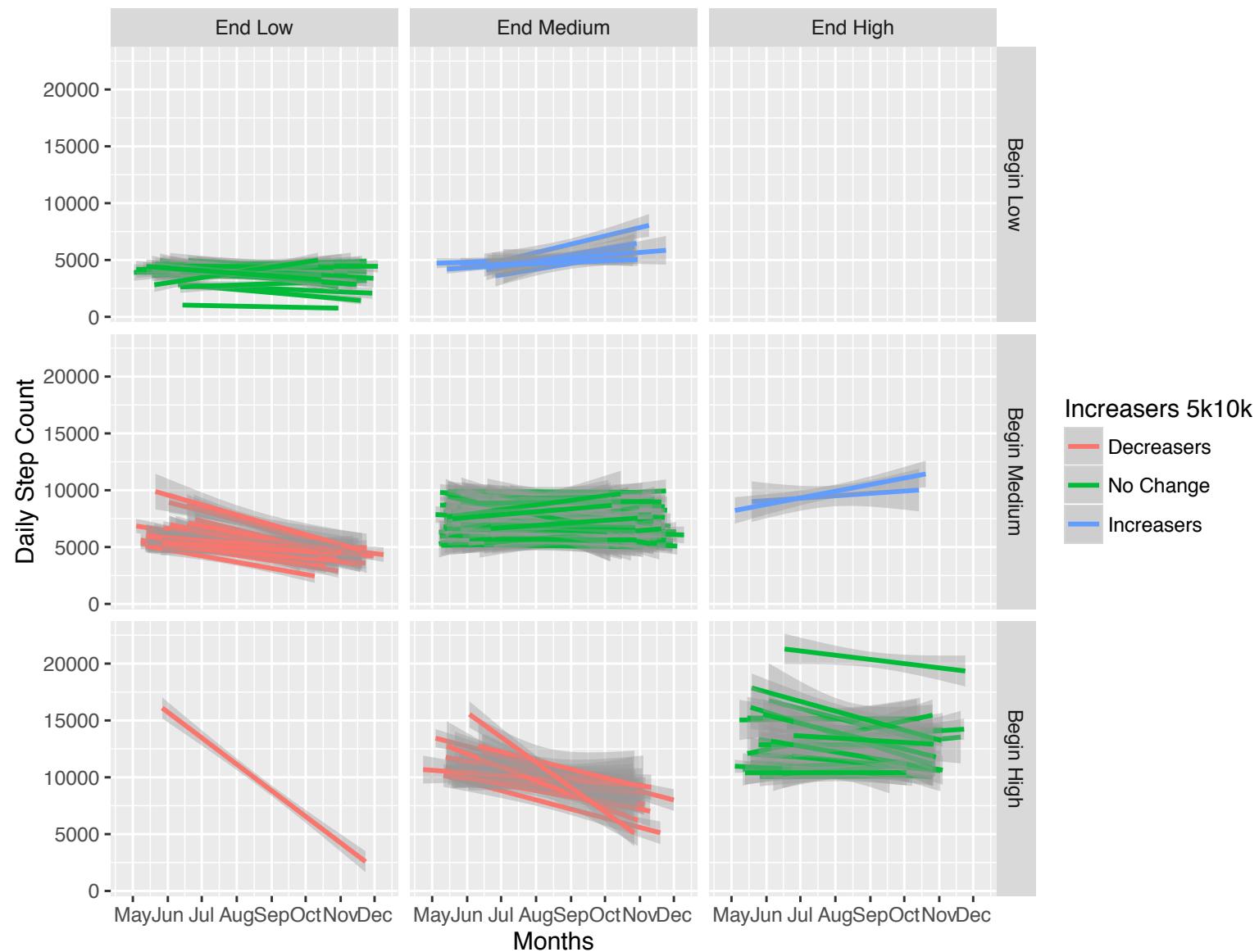
id_10days_calories_bodystates <-
  ggplot(id.week, aes(x=minute_time.utc, y=calories, colour=bodystate)) +
  geom_point(size=1.1) + xlab("Time") + ylab("Calories per minute") +
  scale_colour_manual(drop=TRUE, limits = levels(id.v2$bodystate), values =cols.bd, name="Body states") +
  theme(axis.text.x = element_blank())

id_10days_steps_bodystates <-
  ggplot(id.week, aes(x=minute_time.utc, y=steps, colour=bodystate)) +
  geom_point(size=1.1) + xlab("Time") + ylab("Steps per minute") +
  scale_colour_manual(drop=TRUE, limits = levels(id.v2$bodystate), values =cols.bd, name="Body states") +
  theme(axis.text.x = element_blank())
```

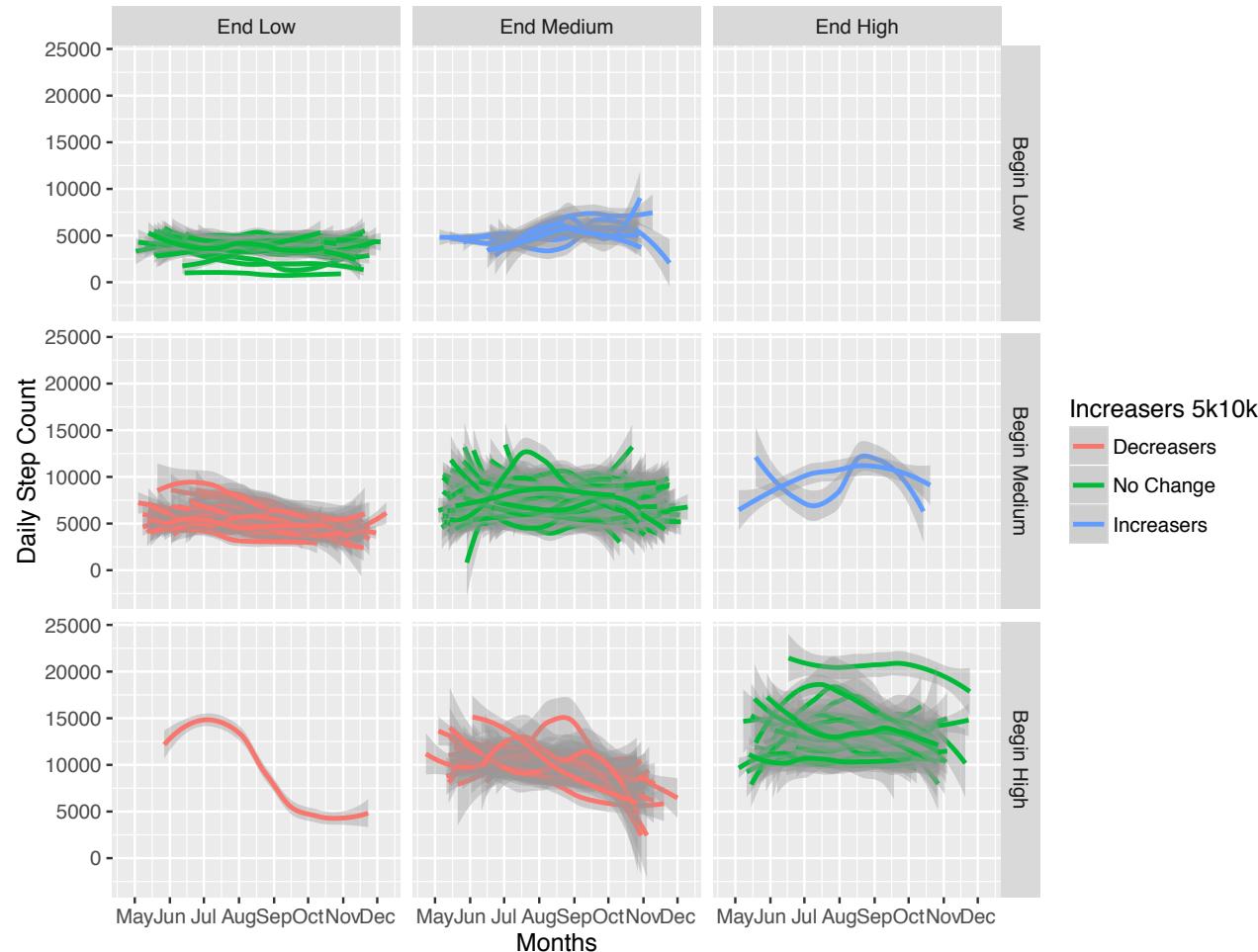
```{r id_10days_bodystates, fig.height=6, fig.width=8, echo=FALSE}
grid.arrange(id_10days_HR_bodystates, id_10days_steps_bodystates, id_10days_calories_bodystates, nrow=3)
```

```

# Visualizing trends in steps/day

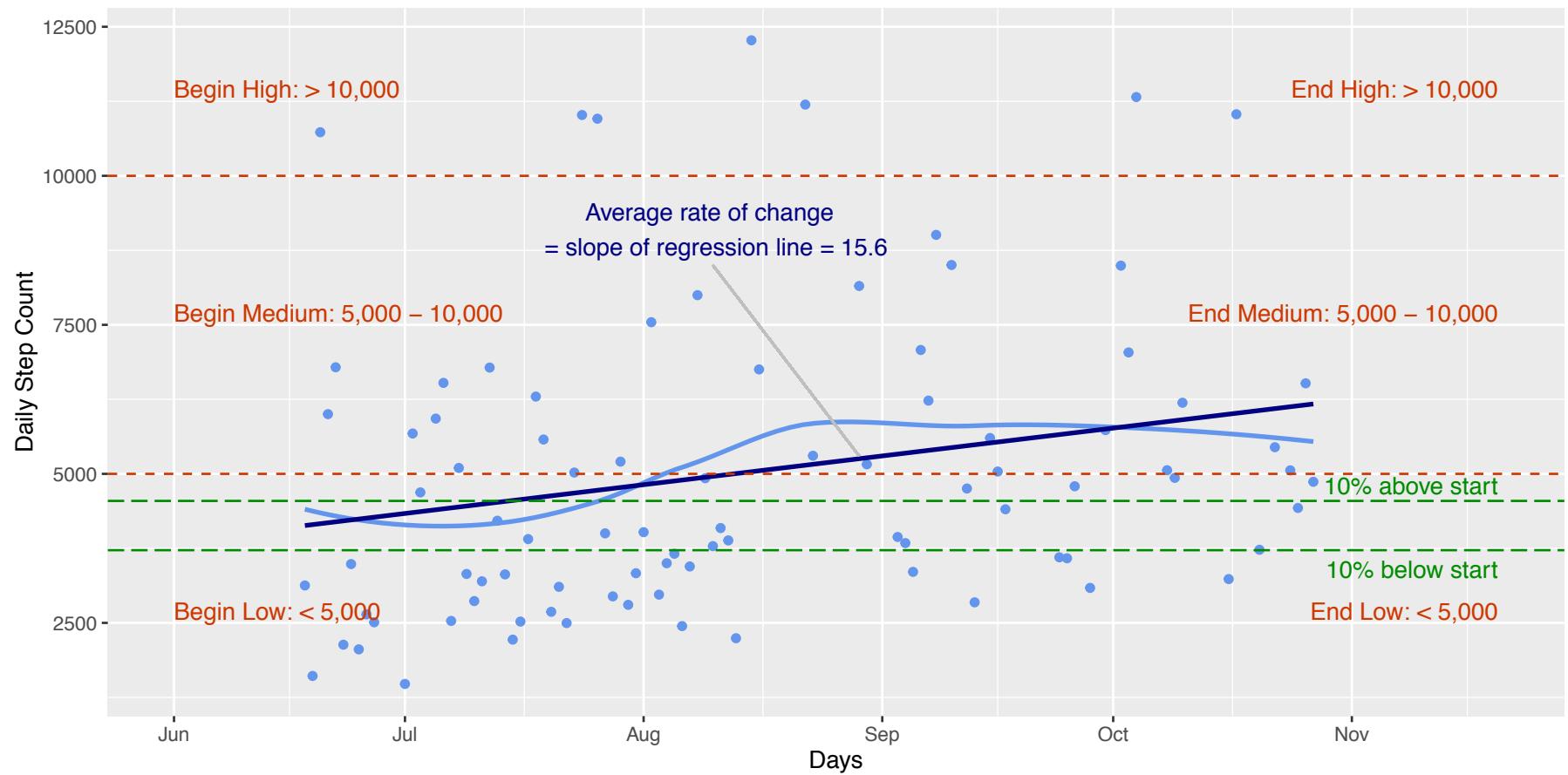


# Visualizing trends in steps/day



```
ggplot(data, aes(x=date, y=steps_SUM, color=resolver.5k10k, group=id))+
  stat_smooth() + xlab("Months") + ylab("Daily Step Count") +
  scale_colour_discrete(name = "5K/10K Trend") +
  facet_grid(steps.t0 ~ steps.t6)
```

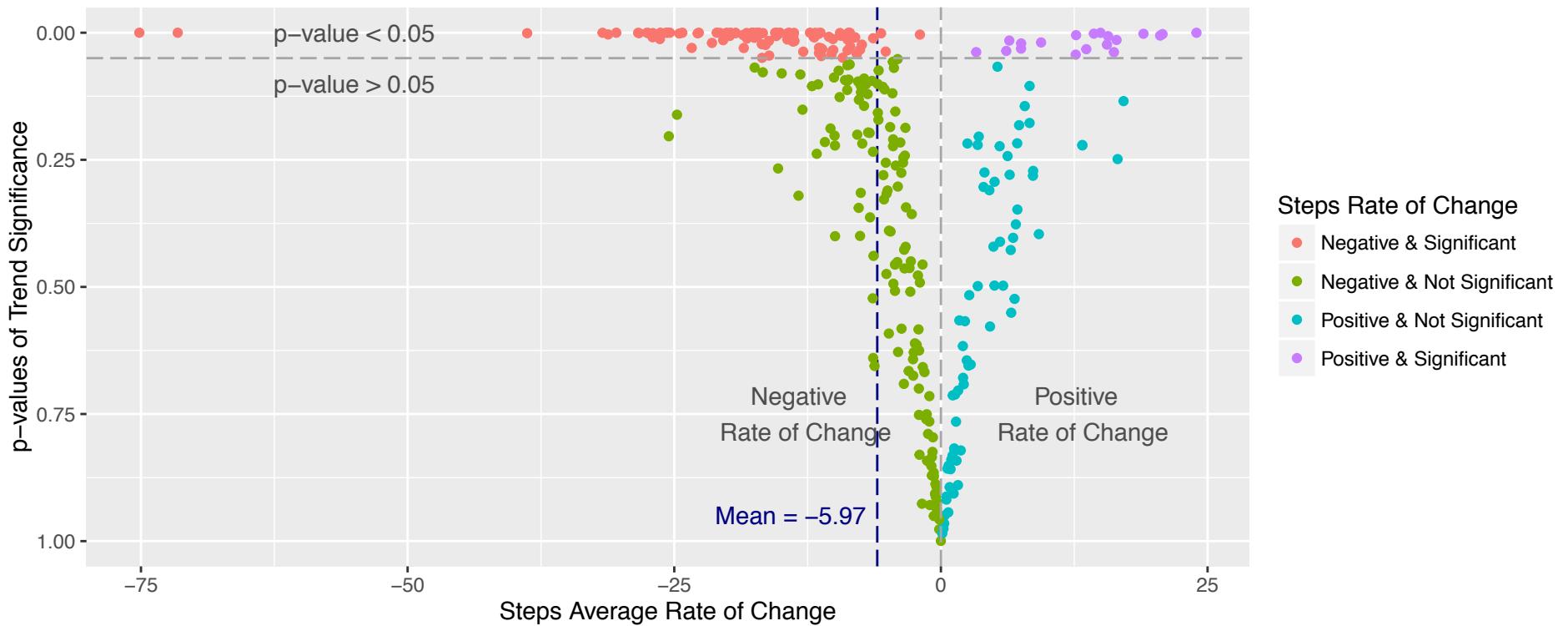
# Visualizing trends in steps/day



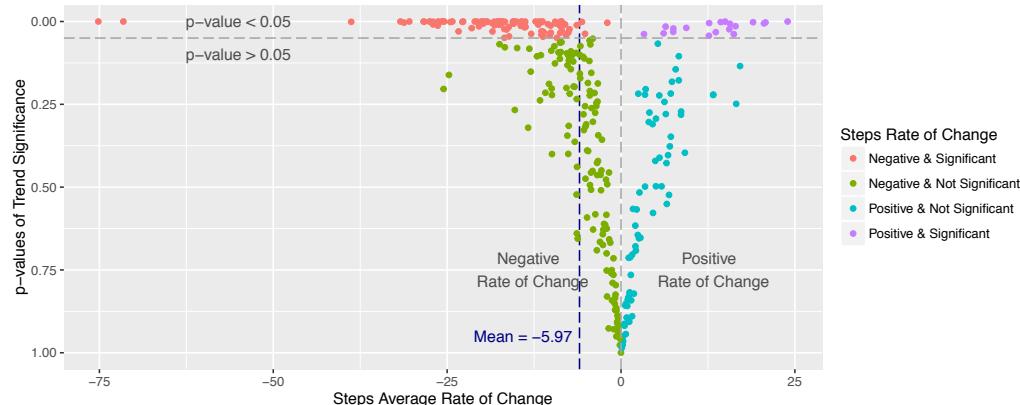
# Visualizing trends in steps/day: Code

```
ggplot(data, aes(x=msr.date, y=steps_SUM)) +  
  geom_point(color="cornflowerblue") +  
  xlab("Days") + ylab("Daily Step Count") +  
  stat_smooth(method="loess", color="cornflowerblue", se=F) +  
  stat_smooth(method = "lm", color="navy", se=F) +  
  geom_hline(yintercept = 1.1*id_intercept, color="green4", linetype = "longdash") +  
  geom_hline(yintercept = .9*id_intercept, color="green4", linetype = "longdash") +  
  geom_hline(yintercept = 5000, color="orangered3", linetype = "dashed") +  
  geom_hline(yintercept = 10000, color="orangered3", linetype = "dashed") +  
# 10% labels  
  annotate("text", x=as.POSIXct("2015-11-20", format = "%Y-%m-%d", tz = "UTC"), y=3400,  
          label="10% below start", color="green4", hjust=1) +  
  annotate("text", x=as.POSIXct("2015-11-20", format = "%Y-%m-%d", tz = "UTC"), y=4800,  
          label="10% above start", color="green4", hjust=1) +  
# begin  
  annotate("text", x=as.POSIXct("2015-06-01", format = "%Y-%m-%d", tz = "UTC"),  
          y=11450, label="Begin High: > 10,000", color="orangered3", hjust=0) +  
  annotate("text", x=as.POSIXct("2015-06-01", format = "%Y-%m-%d", tz = "UTC"),  
          y=7700, label="Begin Medium: 5,000 - 10,000", color="orangered3", hjust=0) +  
  annotate("text", x=as.POSIXct("2015-06-01", format = "%Y-%m-%d", tz = "UTC"),  
          y=2700, label="Begin Low: < 5,000", color="orangered3", hjust=0) +  
# End    #2015-09-10, hjust=0  
  annotate("text", x=as.POSIXct("2015-11-20", format = "%Y-%m-%d", tz = "UTC"),  
          y=11450, label="End High: > 10,000", color="orangered3", hjust=1) +  
  annotate("text", x=as.POSIXct("2015-11-20", format = "%Y-%m-%d", tz = "UTC"),  
          y=7700, label="End Medium: 5,000 - 10,000", color="orangered3", hjust=1) +  
  annotate("text", x=as.POSIXct("2015-11-20", format = "%Y-%m-%d", tz = "UTC"),  
          y=2700, label="End Low: < 5,000", color="orangered3", hjust=1) +  
  geom_segment(aes(x = as.POSIXct("2015-08-10", format = "%Y-%m-%d", tz = "UTC"), y = 8500,  
                    xend = as.POSIXct("2015-08-29", format = "%Y-%m-%d", tz = "UTC"), yend = 5300),  
                color = "grey75") +  
  annotate("text", x=as.POSIXct("2015-08-10", format = "%Y-%m-%d", tz = "UTC"),  
          y=9100, label="Average rate of change \n = slope of regression line = 15.6",  
          color="navy")
```

# Are the slopes significant?

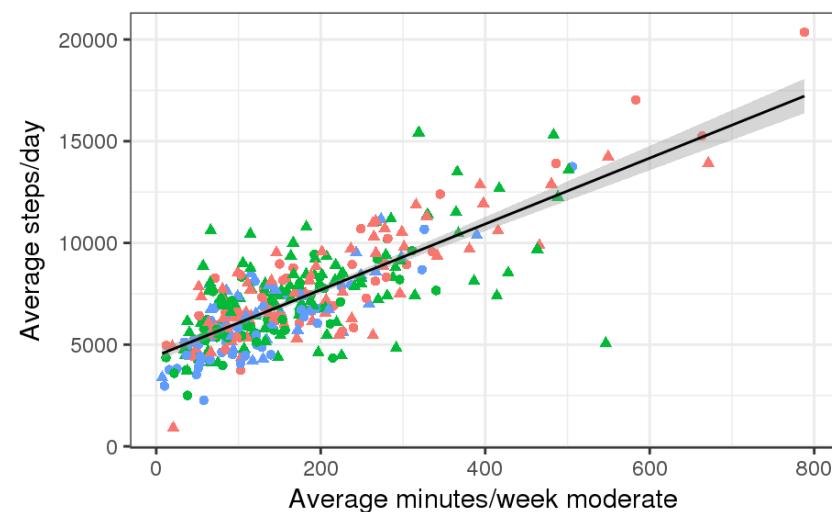
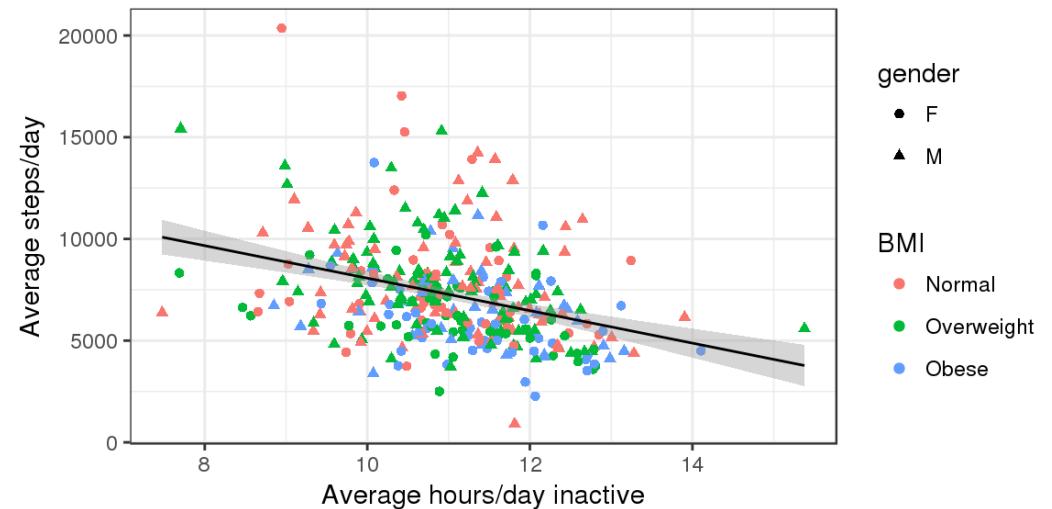
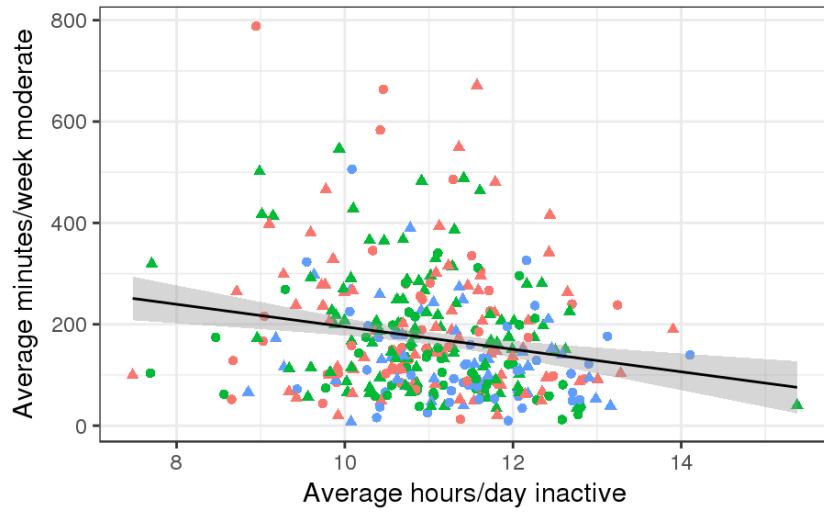


# Are the slopes significant? Code

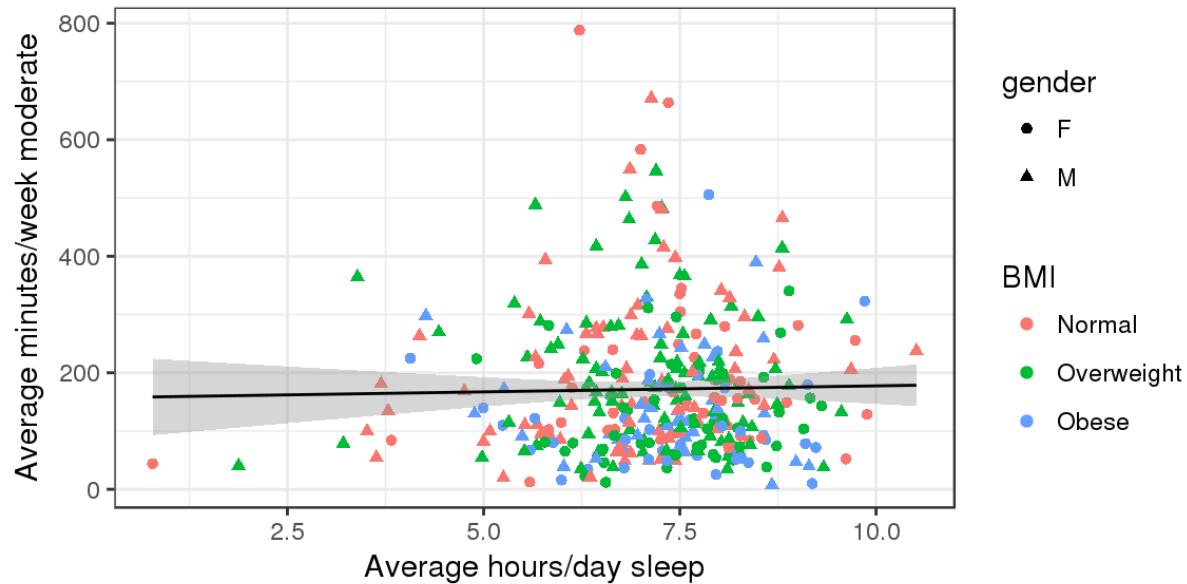


```
ggplot(data, aes(x=slope.steps, y=pv.steps, colour=slope.steps.pv)) +  
  geom_vline(xintercept = mean(data$slope.steps), color="navy", linetype = "longdash") +  
  geom_point() + scale_y_reverse() +  
  xlab("Steps Average Rate of Change") + ylab("p-values of Trend Significance") +  
  geom_hline(yintercept = .05, color="grey65", linetype = "longdash") +  
  geom_vline(xintercept = 0, color="grey65", linetype = "longdash") +  
  annotate("text", x=-55, y=0, label="p-value < 0.05", color="grey30") +  
  annotate("text", x=-55, y=.10, label="p-value > 0.05", color="grey30") +  
  annotate("text", x=-13, y=.75, label="Negative \n Rate of Change", color="grey30") +  
  annotate("text", x=13, y=.75, label="Positive \n Rate of Change", color="grey30") +  
  annotate("text", x=-7, y=.95, label="Mean = -5.97", color="navy", hjust=1) +  
# Set legend  
  scale_color_discrete(name = "Steps Rate of Change",  
    labels = c("Negative & Significant", "Negative & Not Significant",  
              "Positive & Not Significant", "Positive & Significant"))
```

# Associations between steps, moderate activity, & inactivity



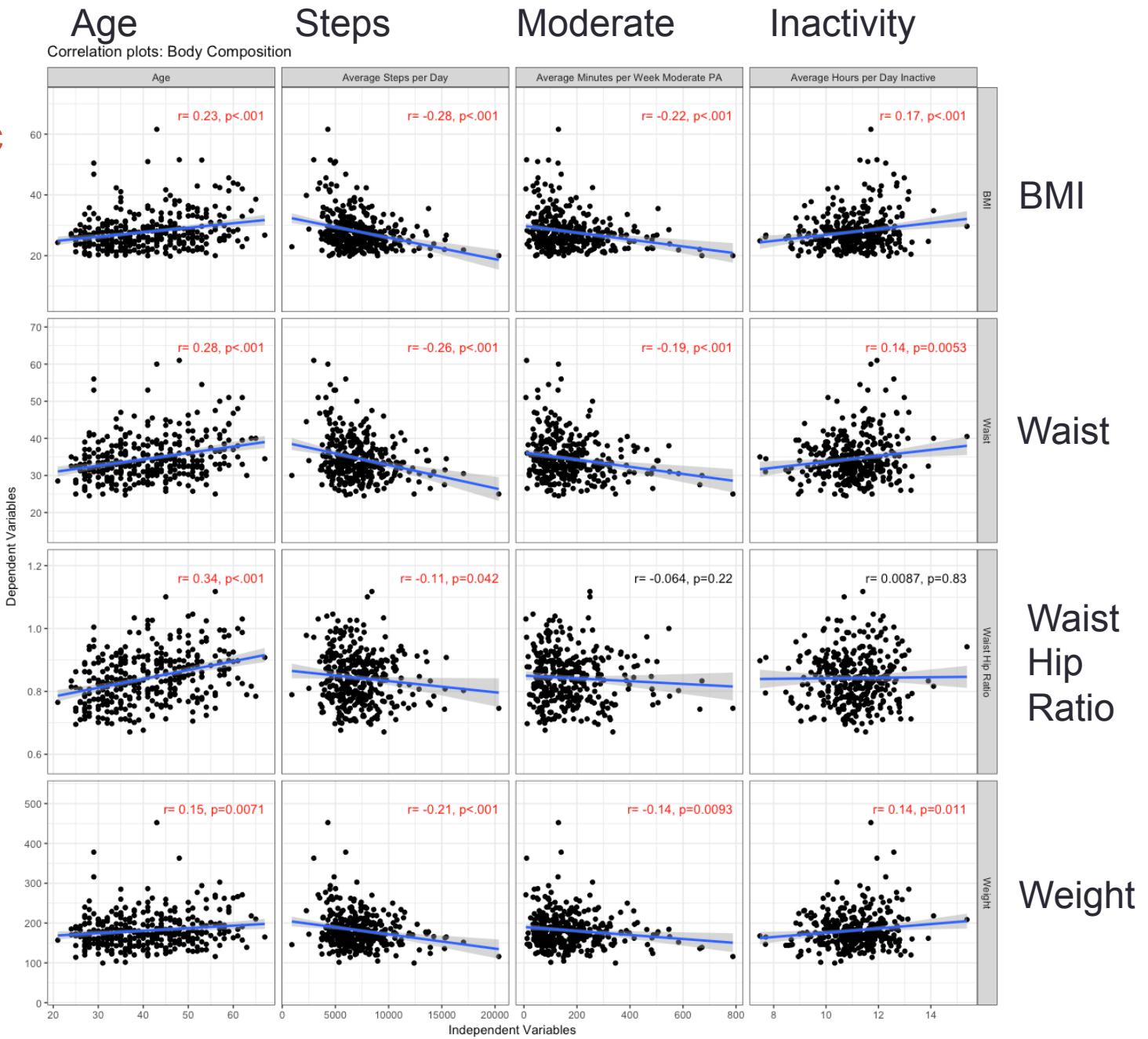
# Associations with sleep



```
ggplot(data,
  aes(x=HrsD_sleep_mean, y=ModerateMinWk_mean)) +
  geom_point(aes(color=BMI_T6, shape=gender)) +
  stat_smooth(method = "lm", color="black", size=.5) +
  xlab("Average hours/day sleep") +
  ylab("Average minutes/week moderate") +
  scale_colour_discrete(name = "BMI") +
  theme_bw()
```

Cardiometabolic  
risk factors  
vs.  
age & activity

Made data  
“doubly long”  
to facet  
two sets of  
variables



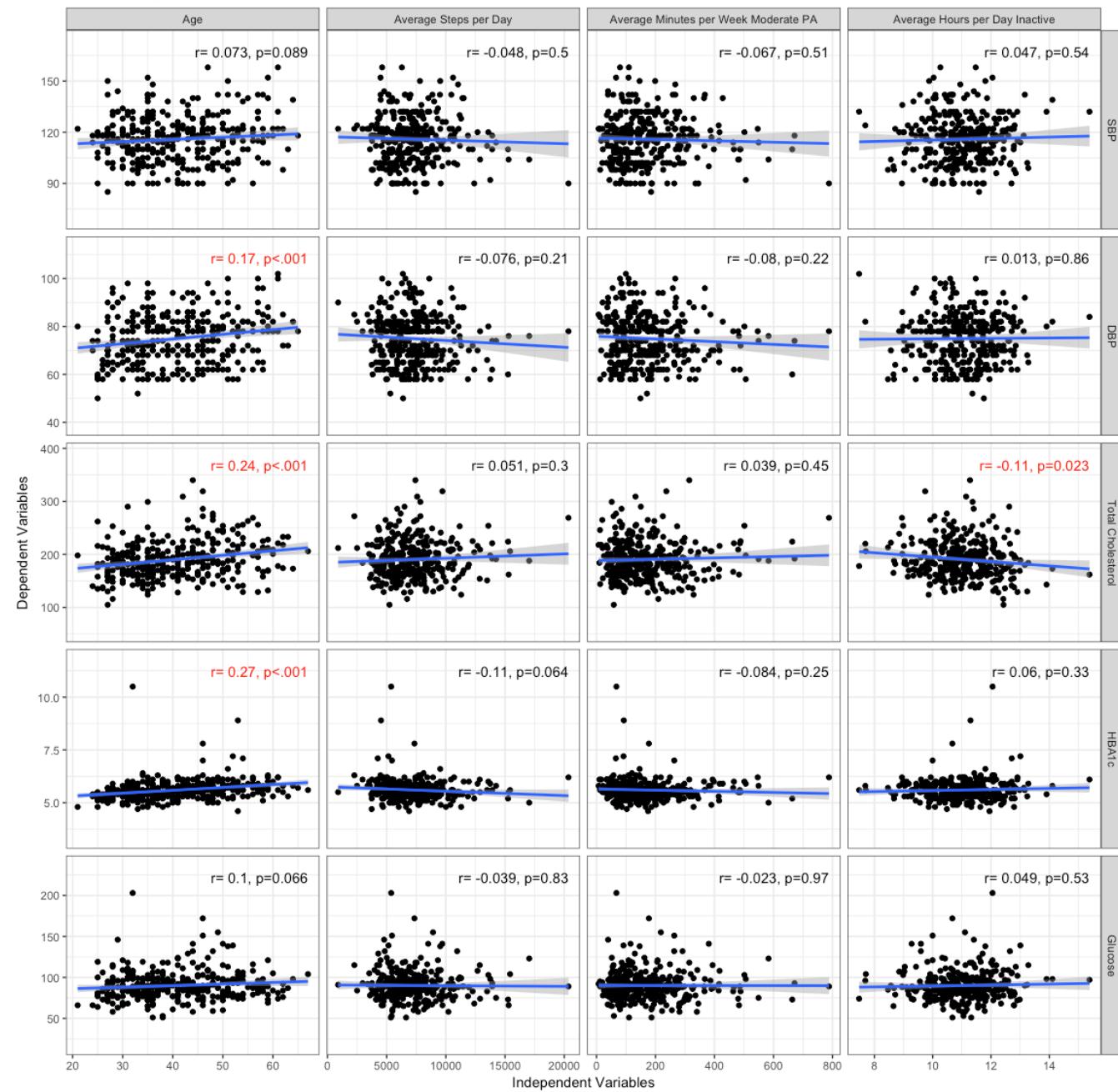
## Age

## Steps

## Moderate

## Inactivity

Correlation plots: Clinical



SBP

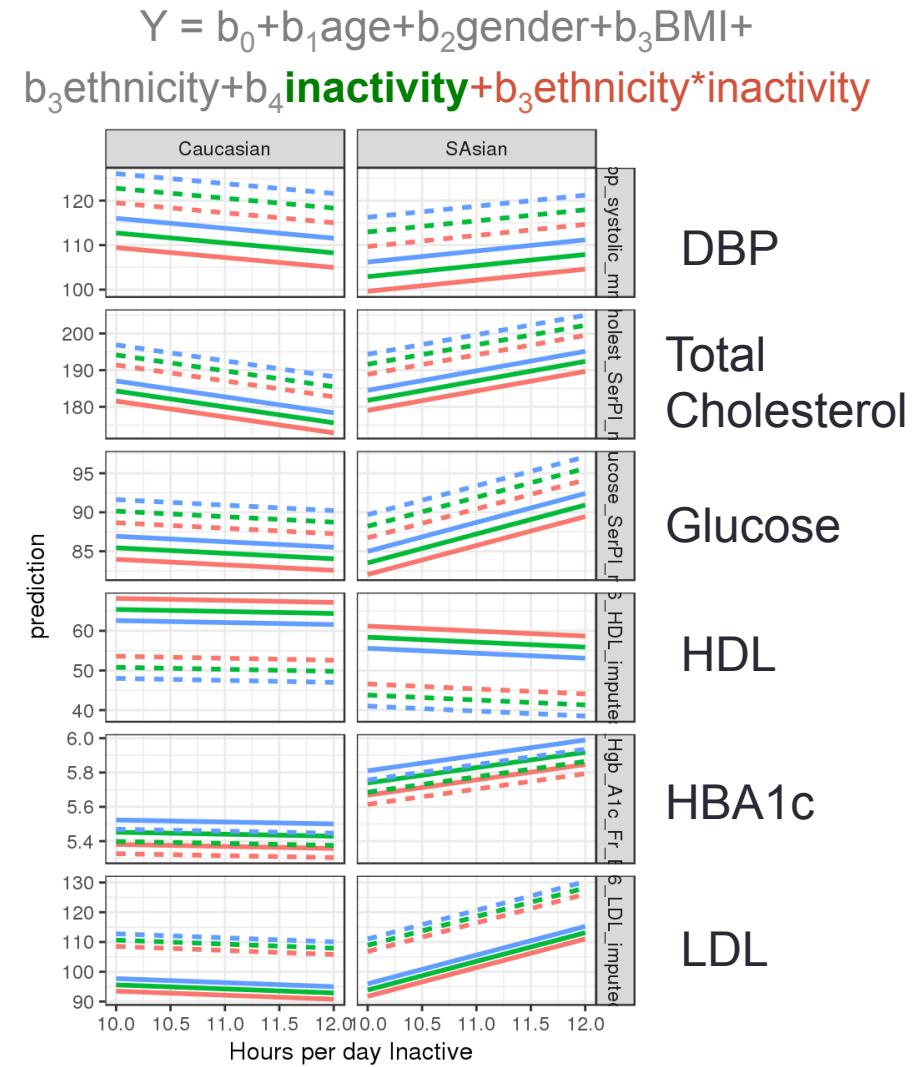
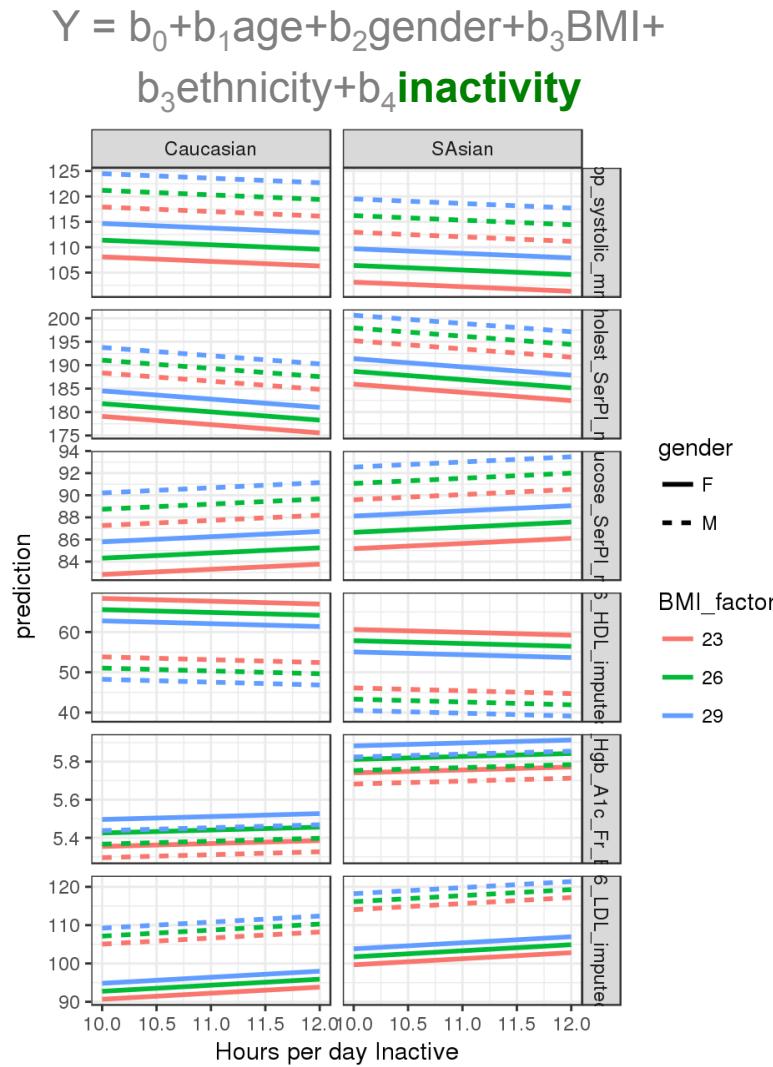
DBP

Total Cholesterol

HbA1c

Glucose

# Visualizing & comparing regression models



# Code for plotting regression models (1)

```
Clinical_long <- gather(Clinical_wide, variable, measurement,  
t6_bp_systolic_mmhg:t6_LDL_imputed)  
Clinical_long$variable <- as.factor(Clinical_long$variable)  
  
Clinical_2xlong_PA <- gather(Clinical_long, PAtype, PValue,  
stepsD_mean_c:HrsD_inactive_mean_c)  
Clinical_2xlong_PA$PAtype <- as.factor(Clinical_2xlong_PA$PAtype)
```

```
nest_Clinical_2xlong_PA <- Clinical_2xlong_PA %>%  
  group_by(PAtype, variable) %>%  
  nest()
```

```
> nest_Clinical_2xlong_PA  
# A tibble: 27 x 3  
      PAtype          variable        data  
      <fctr>          <fctr>        <list>  
 1 stepsD_mean_c  t6_bp_systolic_mmhg <tibble [203 x 7]>  
 2 stepsD_mean_c  t6_bp_diastolic_mmhg <tibble [203 x 7]>  
 3 stepsD_mean_c    t6_Hgb_A1c_Fr_Bld <tibble [203 x 7]>  
 4 stepsD_mean_c   t6_Glucose_SerPl_mCnc <tibble [203 x 7]>  
 5 stepsD_mean_c   t6_Cholest_SerPl_mCnc <tibble [203 x 7]>  
 6 stepsD_mean_c   t6_CholRatio_imputed <tibble [203 x 7]>  
 7 stepsD_mean_c       t6_HDL_imputed <tibble [203 x 7]>  
 8 stepsD_mean_c       t6_nonHDL_imputed <tibble [203 x 7]>  
 9 stepsD_mean_c       t6_LDL_imputed <tibble [203 x 7]>  
10 ModerateMinWk_mean_c  t6_bp_systolic_mmhg <tibble [203 x 7]>  
# ... with 17 more rows
```

# Code for plotting regression models (2)

```
Fcn_LM_rhsPA_BMI <- function(df) {  
  lm(measurement ~ gender + age_c + BMI_c + race + PAvalue, data=df)  
}  
  
Fcn_LM_rhsPA_SAvsPA <- function(df) {  
  lm(measurement ~ gender +age_c +BMI_c +race +PAvalue +race*PAvalue, data=df)  
}  
  
nest_Clinical_2xlong_PA <- nest_Clinical_2xlong_PA %>%  
  mutate(out_LM_BMI = map(data, Fcn_LM_rhsPA_BMI)) %>%  
  mutate(out_LM_SAvsPA = map(data, Fcn_LM_rhsPA_SAvsPA))
```

```
# A tibble: 27 x 5  
  PAtype      variable      data  out_LM_BMI out_LM_SAvsPA  
  <fctr>       <fctr>       <list>   <list>       <list>  
 1 stepsD_mean_c t6_bp_systolic_mmhg <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 2 stepsD_mean_c t6_bp_diastolic_mmhg <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 3 stepsD_mean_c t6_Hgb_A1c_Fr_Bld <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 4 stepsD_mean_c t6_Glucose_SerPl_mCnc <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 5 stepsD_mean_c t6_Cholest_SerPl_mCnc <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 6 stepsD_mean_c t6_CholRatio_imputed <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 7 stepsD_mean_c t6_HDL_imputed <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 8 stepsD_mean_c t6_nonHDL_imputed <tibble [203 x 7]> <S3: lm>   <S3: lm>  
 9 stepsD_mean_c t6_LDl_imputed <tibble [203 x 7]> <S3: lm>   <S3: lm>  
10 ModerateMinWk_mean_c t6_bp_systolic_mmhg <tibble [203 x 7]> <S3: lm>   <S3: lm>  
# ... with 17 more rows
```

# Code for plotting regression models (3)

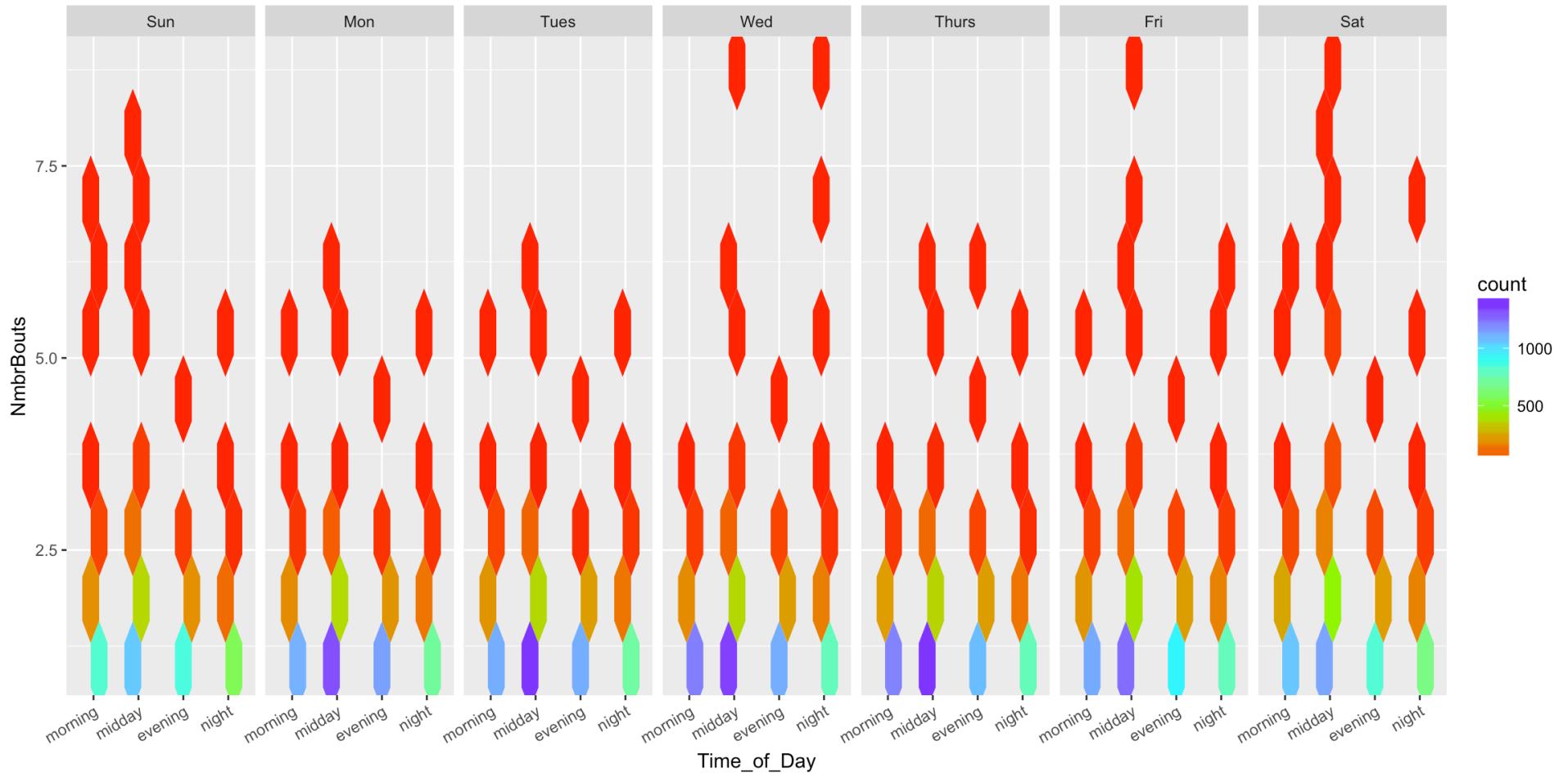
```
LM_predict_HrsD_inactive <- function(LM){  
  # need to create data with values to predict  
  temp <- predict_grid_HrsD_inactive  
  temp$prediction <- predict(LM, newdata=predict_grid_HrsD_inactive)  
  temp  
}  
  
nest_HrsD_inactive <- nest_Clinical_2xlong_PA %>%  
  filter(PAtype == "HrsD_inactive_mean_c") %>%  
  mutate(predict_LM_BMI_HrsD_inactive =  
         map(out_LM_BMI, LM_predict_HrsD_inactive)) %>%  
  mutate(predict_SAvsPA_HrsD_inactive =  
         map(out_LM_SAvsPA, LM_predict_HrsD_inactive))  
  
# Ready to unnest:  
predict_SAvsPA_HrsD_inactive_unnest <-  
  unnest(nest_HrsD_inactive, predict_SAvsPA_HrsD_inactive)
```

# Code for plotting regression models (4)

```
> predict_SAvsPA_HrsD_inactive_unnest
# A tibble: 324 × 12
#>   PAtype      variable age_c     race    PAvalue    BMI_c gender HrsD_inactive_uncntrd
#>   <fctr>      <fctr>  <dbl>    <fctr>    <dbl>    <dbl> <fctr>          <dbl>
#> 1 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 Caucasian -1.08643815 -4.21622    F           10
#> 2 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 SAsian   -1.08643815 -4.21622    F           10
#> 3 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 Caucasian -0.08643815 -4.21622    F           11
#> 4 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 SAsian   -0.08643815 -4.21622    F           11
#> 5 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 Caucasian  0.91356185 -4.21622    F           12
#> 6 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 SAsian   0.91356185 -4.21622    F           12
#> 7 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 Caucasian -1.08643815 -1.21622    F           10
#> 8 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 SAsian   -1.08643815 -1.21622    F           10
#> 9 HrsD_inactive_mean_c t6_bp_systolic_mmhg     0 Caucasian -0.08643815 -1.21622    F           11
#> 10 HrsD_inactive_mean_c t6_bp_systolic_mmhg    0 SAsian   -0.08643815 -1.21622    F           11
# ... with 314 more rows, and 4 more variables: HrsD_inactive_factor <fctr>, BMI_uncntrd <dbl>, BMI_factor <fctr>,
#   prediction <dbl>
```

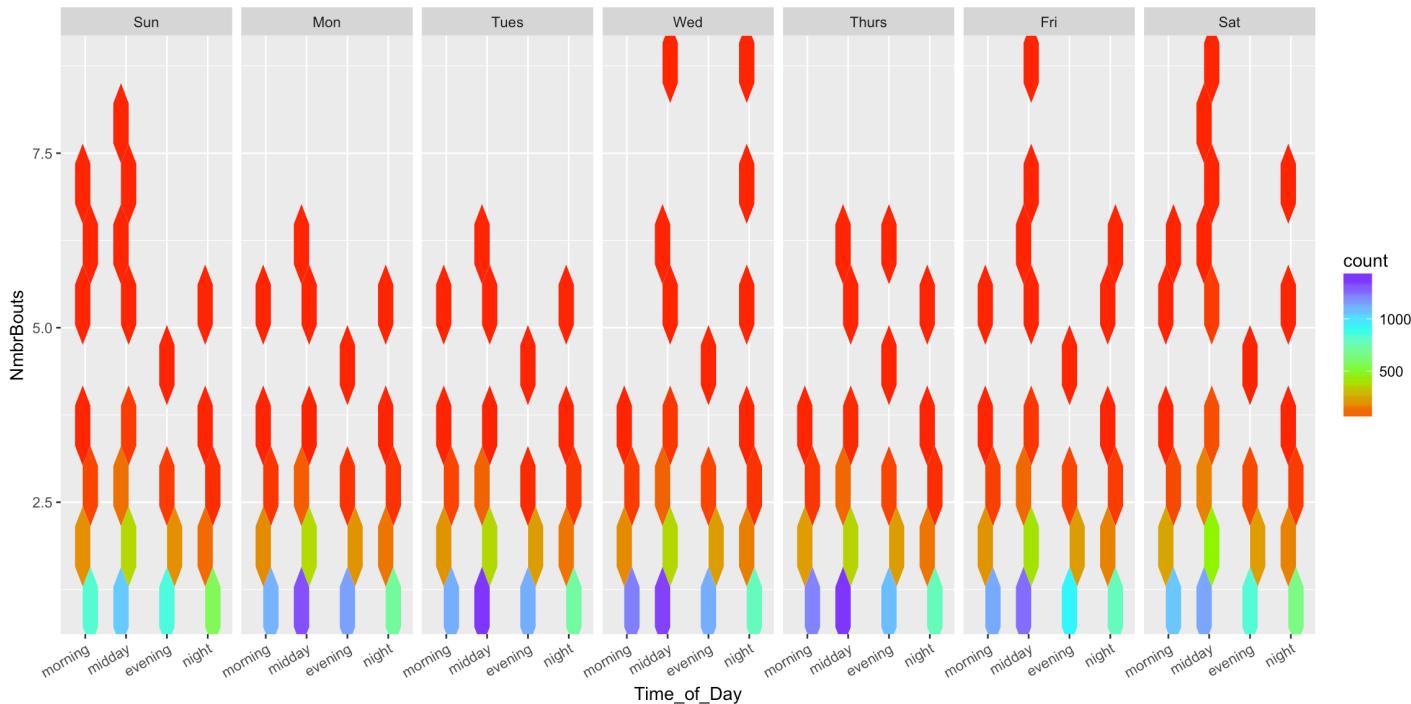
```
ggplot(predict_SAvsPA_HrsD_inactive_unnest,
       aes(x=HrsD_inactive_uncntrd, y=prediction,
            color = BMI_factor, linetype = gender)) +
  geom_line(size = 1) +
  facet_grid(variable ~ race, scales = "free") +
  xlab("Hours per day Inactive") +
  theme_bw()
```

# Bouts of moderate activity



Counts are based on all bouts from all participants

# Bouts of moderate activity: Hex plot code



```
ggplot(data, aes(x=Time_of_Day, y=NmbrBouts)) +  
  facet_grid(~ day_of_week) +  
  geom_hex(bins=2*4) +  
  scale_fill_gradientn(colours = rainbow(4)) +  
  theme(axis.text.x=element_text(angle=30, hjust=1))
```

# Thank you!

## OHSU collaborators

- Knight Cardiovascular Institute
  - Luke Burchill
  - Yuliang Wang
  - Jessica Minnier
  - Allen Rassa
  - Vaughan Tuohy
- Health Promotion & Sports Medicine
  - Beth Smith
- OHSU-PSU School of Public Health
  - Jodi Lapidus
  - Nora Fino
  - David Yanez
- Oregon Institute of Occupational Health Sciences
  - Steven Shea
  - Saurabh Thosar