PoA_GraphComparisons

2023-08-02

Libraries

Read in Data

```
source("~/Drexel/PHACS:AMP/Code/root.R")

chang <- read.csv(paste0(root, "Exported_Data/chang.csv"))
belsky <- read.csv(paste0(root, "Exported_Data/belsky.csv"))
kdm_poa <- read.csv(paste0(root, "Exported_Data/kdm.csv"))

kdm <- read.csv(paste0(root, "Exported_Data/kdm_all_comp.csv"))
```

Get Individual slopes of KDM

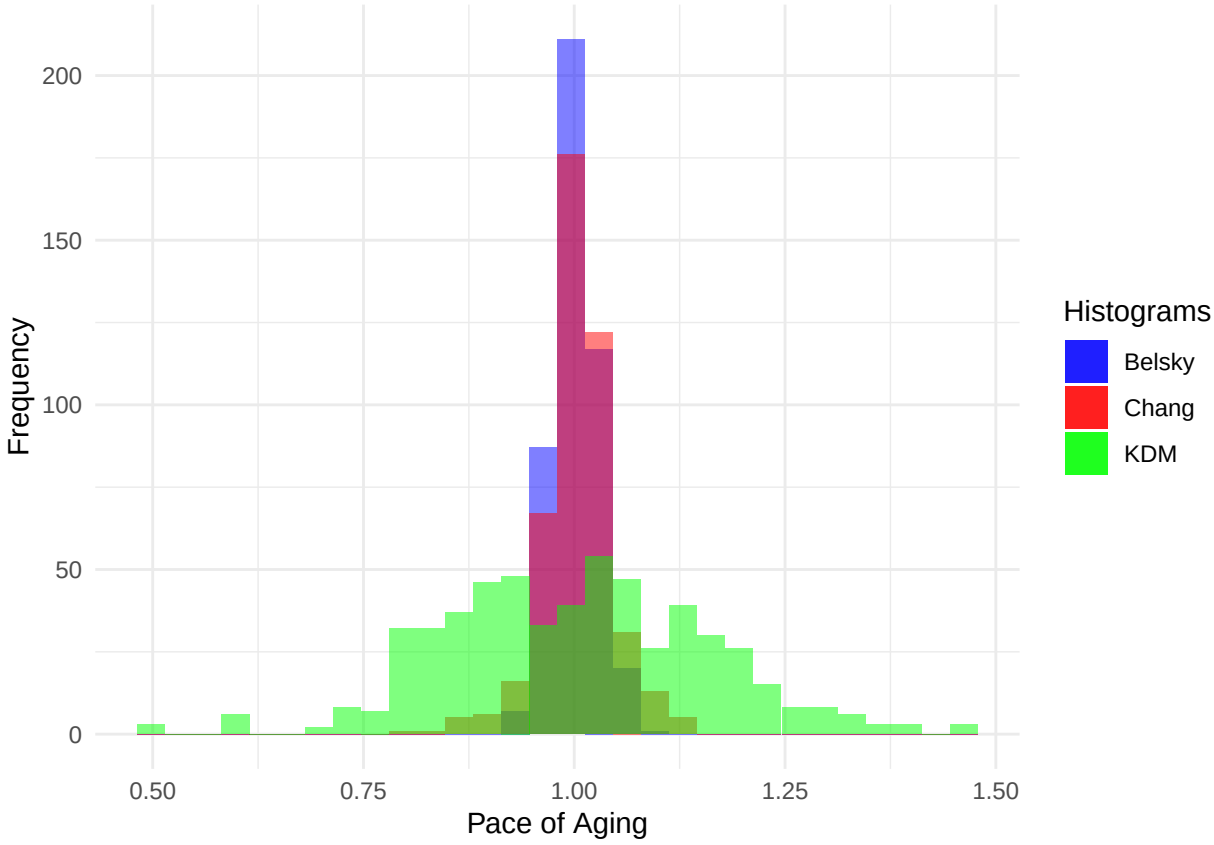
```
# # Step 1: Prepare the data (if not already done)
# # Data should be in the format: participantID, timepoint, kdm_biological_age
# kdm_prep <- kdm %>% select(publicID, week_scale, scale_base_age, kdm, age, scaled_age)
#
# # Step 2: Fit a linear regression model for each individual
# kdm_slope_data <- kdm_prep %>%
#   group_by(publicID) %>%
#   summarize(slope = coef(lm(kdm ~ week_scale*scale_base_age))["week_scale"])
#
# hist(kdm_slope_data$slope*0.92)
#
#
# # outliers_list <- kdm_slope_data %>% filter(abs(slope) > 1.5) %>% select(publicID) %>% as.list
# # kdm_slope_data %>% filter(abs(slope) > 1.5)
# # kdm %>% filter(publicID %in% outliers_list$publicID) %>% arrange(publicID) %>% select(publicID, wee
```

Graph three Histograms of Pace of Aging

Overlayed

```
ggplot(belsky, aes(x = poa, fill = "Belsky")) +
  geom_histogram(alpha = 0.5) +
  geom_histogram(data = chang, aes(x = poa, fill = "Chang"), alpha = 0.5) +
  geom_histogram(data = kdm_poa, aes(x = poa, fill = "KDM"), alpha = 0.5) +
  labs(x = "Pace of Aging", y = "Frequency") +
  scale_fill_manual(name = "Histograms", values = c("blue", "red", "green")) +
  theme_minimal()
```

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



Panels

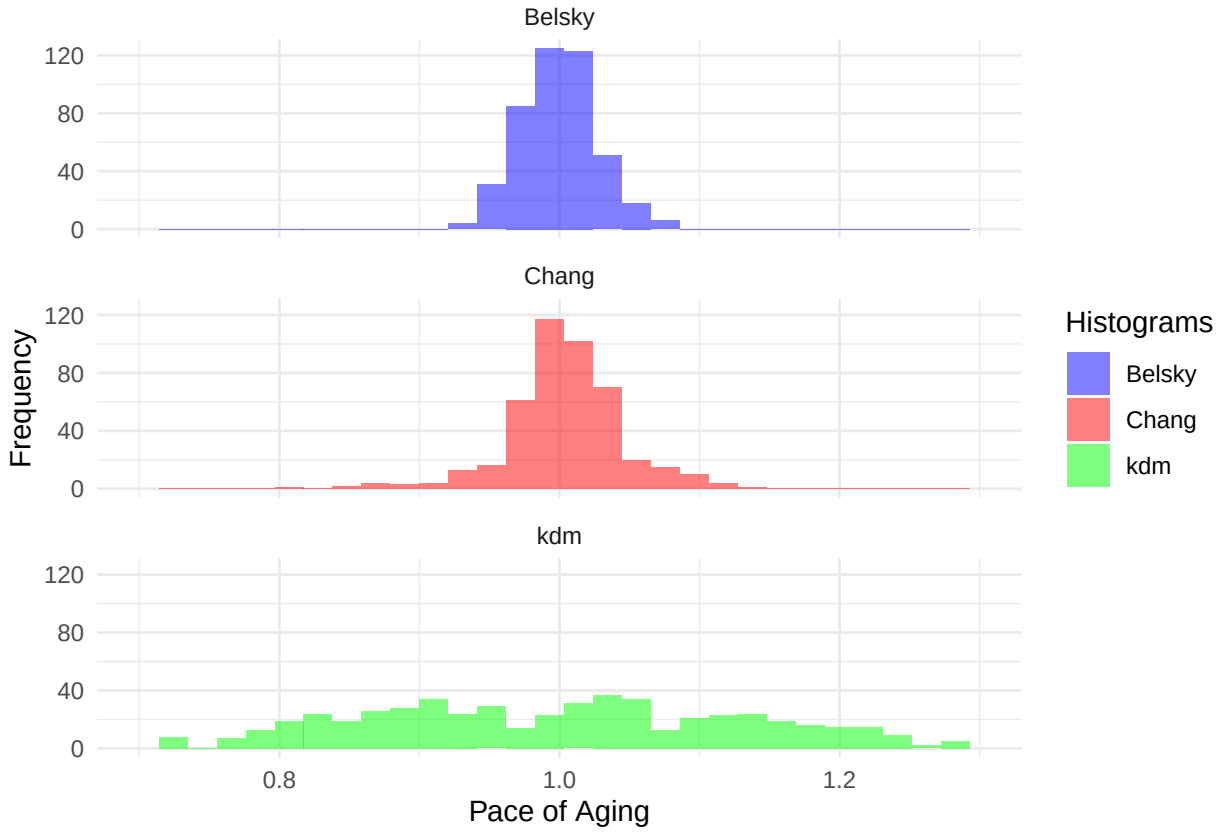
```
panel_graph <- rbind(belsky %>% select(poa, group),
                     chang %>% select(poa, group),
                     kdm_poa %>% select(poa, group))

# Plot the histogram with proportions of poa within each variable
ggplot(data = panel_graph, aes(x = poa, fill = group)) +
  geom_histogram(alpha = 0.5, position = "identity") +
  labs(x = "Pace of Aging", y = "Frequency") +
  scale_fill_manual(name = "Histograms", values = c("Belsky" = "blue", "Chang" = "red", "kdm" = "green"
```

```
  theme_minimal() +
  facet_wrap(~group, ncol = 1) +
  scale_x_continuous(limits = c(0.7, 1.3))
```

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

```
## Warning: Removed 27 rows containing non-finite values (`stat_bin()`).
```

```
## Warning: Removed 6 rows containing missing values (`geom_bar()`).
```

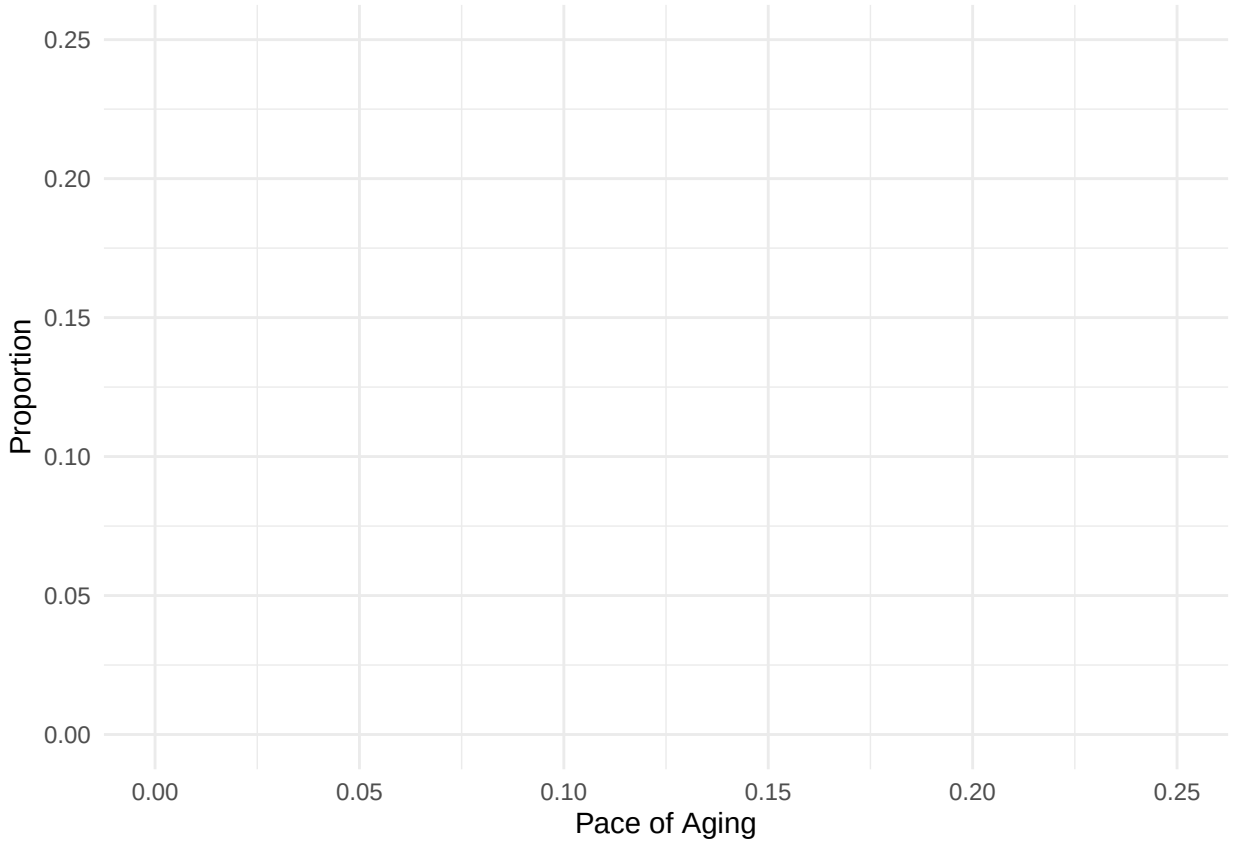



```
ggplot(data = belsky, aes(x = poa, fill = "KDM")) +
  geom_histogram(alpha = 0.5, position = "identity", aes(y = ..count.. / sum(..count..))) +
  labs(x = "Pace of Aging", y = "Proportion") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 0.25)) +
  scale_x_continuous(limits = c(0, 0.25))
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

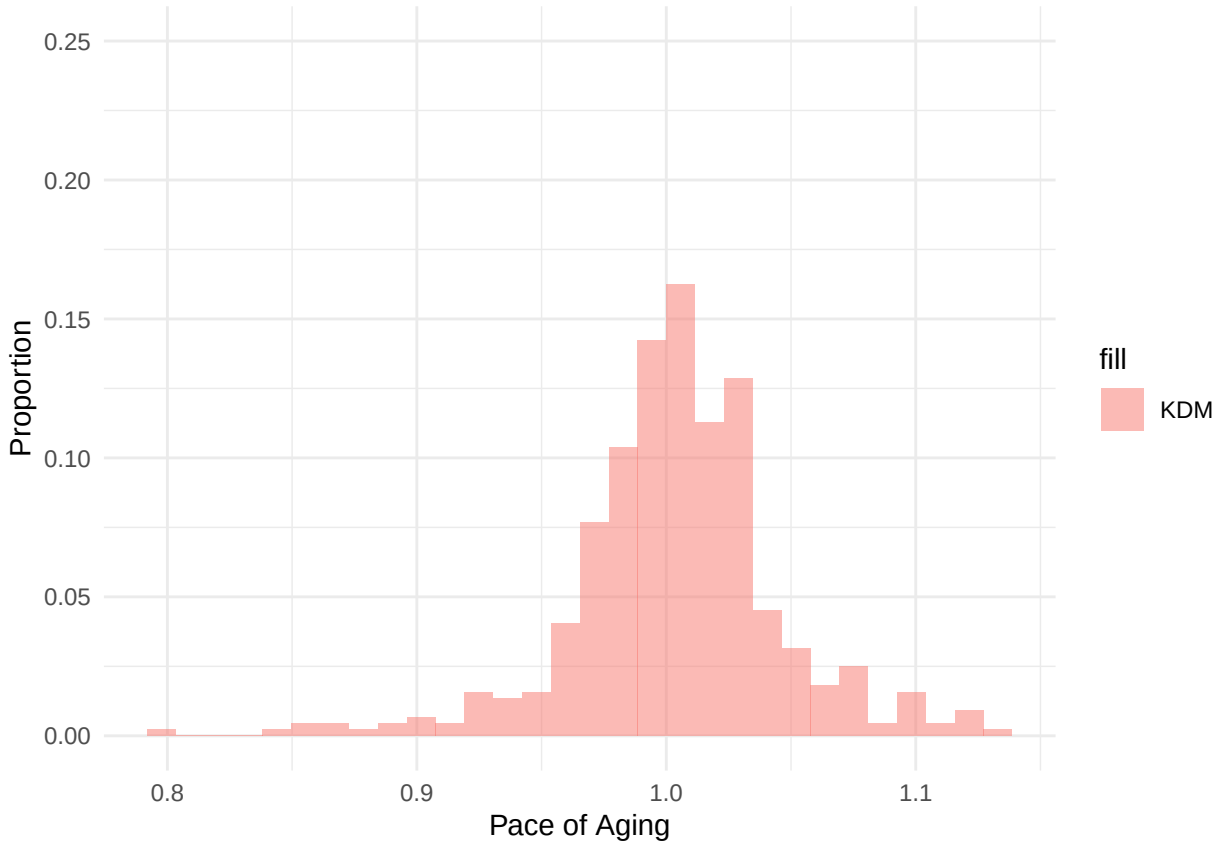
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 443 rows containing non-finite values (`stat_bin()`).
```




```
ggplot(data = chang, aes(x = poa, fill = "KDM")) +
  geom_histogram(alpha = 0.5, position = "identity", aes(y = ..count.. / sum(..count..))) +
  labs(x = "Pace of Aging", y = "Proportion") +
  theme_minimal() +
  scale_y_continuous(limits = c(0, 0.25))
```

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



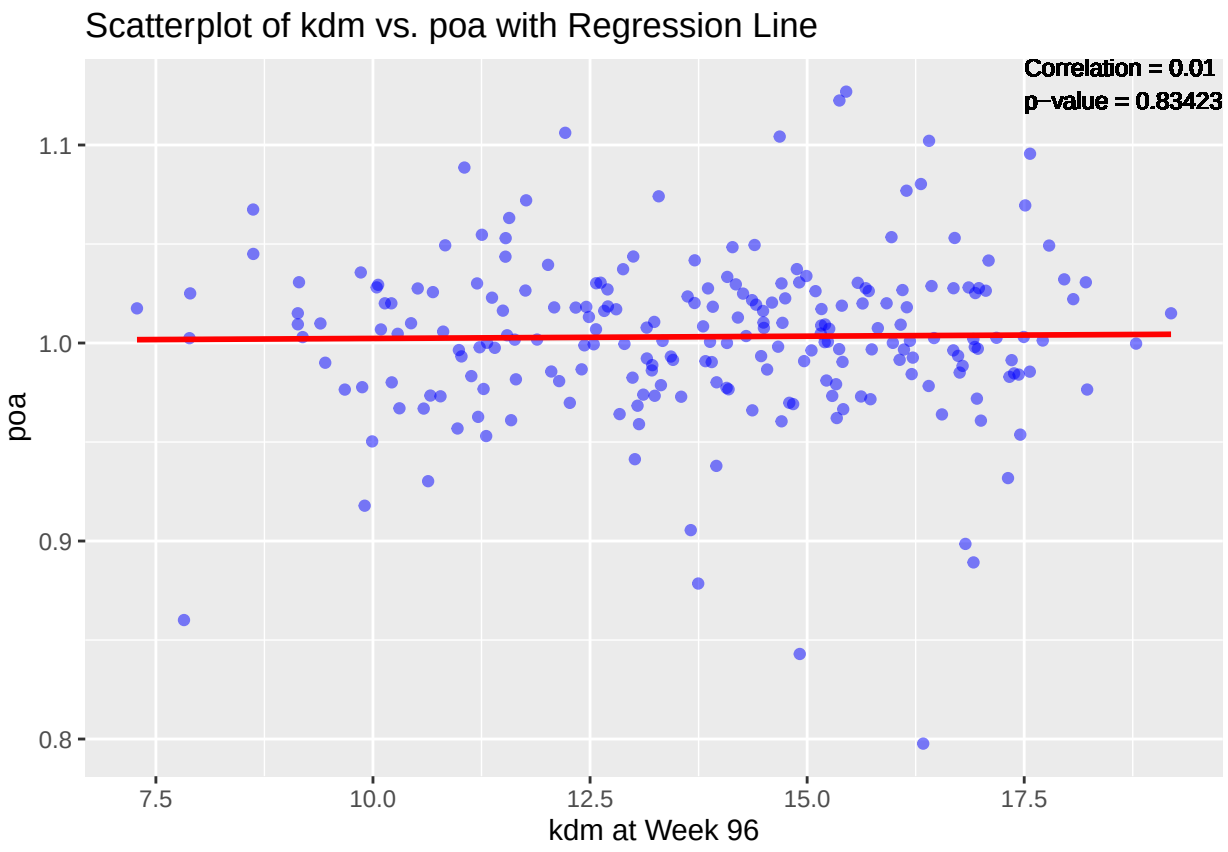

```
# ggplot(data = kdm_slope_data_forgraph, aes(x = poa, fill = "KDM")) +
#   geom_histogram(alpha = 0.5, position = "identity", aes(y = ..count.. / sum(..count..))) +
#   labs(x = "Pace of Aging", y = "Proportion") +
#   theme_minimal() +
#   scale_y_continuous(limits = c(0, 0.25))
```

Pace of Aging vs. KDM at 96 Weeks

```
chang_kdm <- merge(kdm %>% filter(week == 96) %>% select(publicID, kdm),
      chang %>% select(publicID, poa),
      by = "publicID")

chang_kdm_poa <- merge(kdm_poa %>% select(publicID, poa),
      chang %>% select(publicID, poa),
      by = "publicID")

belsky_kdm <- merge(kdm %>% filter(week == 96) %>% select(publicID, kdm),
      belsky %>% select(publicID, poa),
      by = "publicID")

belsky_kdm_poa <- merge(kdm_poa %>% select(publicID, poa),
      belsky %>% select(publicID, poa),
      by = "publicID")
```

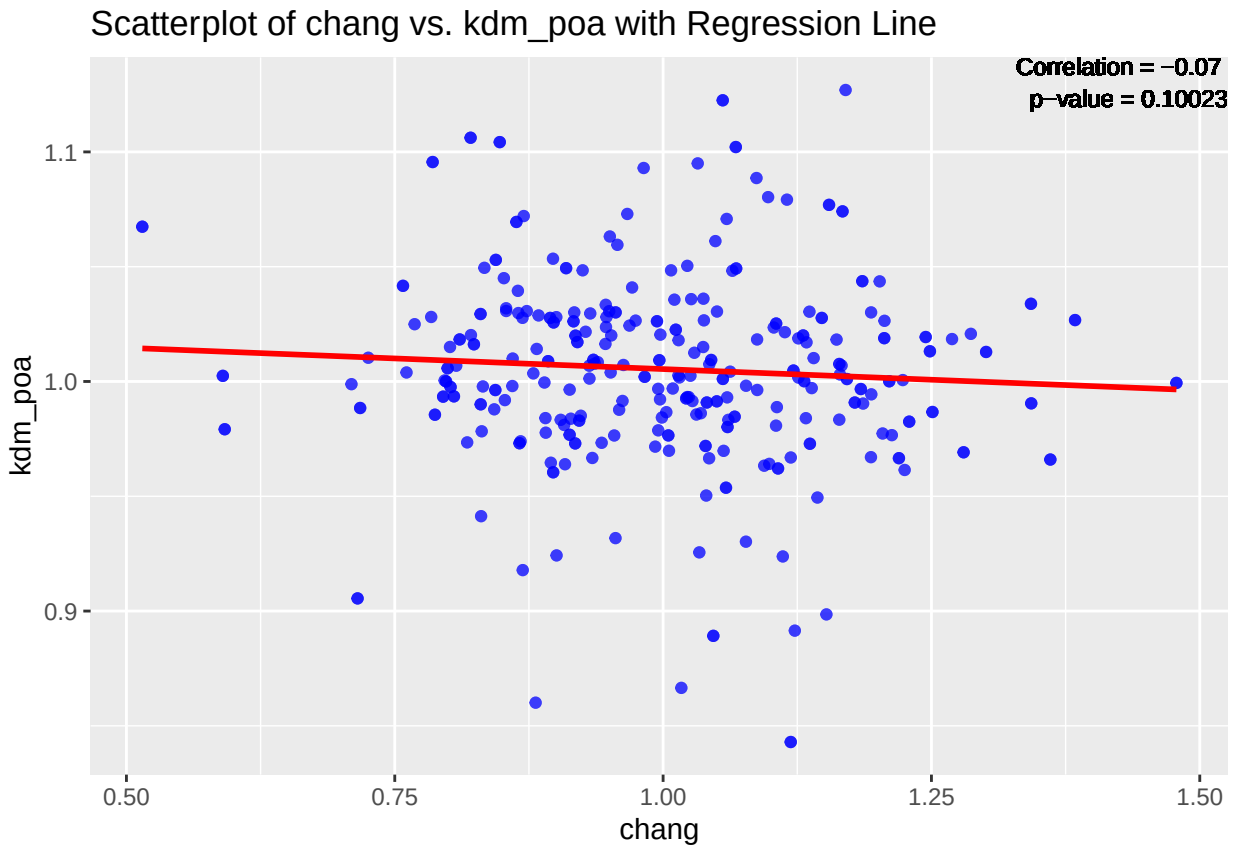
```r
chang_belsky <- merge(chang %>% select(publicID, poa),
      belsky %>% select(publicID, poa),
      by = "publicID")


# Calculate correlation and p-value
correlation_data_chang <- chang_kdm %>%
  select(kdm, poa) %>%
  drop_na()  # Remove rows with missing values

correlation_chang <- cor.test(correlation_data_chang$kdm, correlation_data_chang$poa)

# Calculate correlation and p-value
correlation_data_chang_poa <- chang_kdm_poa %>%
  select(poa.x, poa.y) %>%
  drop_na()  # Remove rows with missing values

correlation_chang_poa <- cor.test(correlation_data_chang_poa$poa.x, correlation_data_chang_poa$poa.y)

# Calculate correlation and p-value
correlation_data_belsky <- belsky_kdm %>%
  select(kdm, poa) %>%
  drop_na()  # Remove rows with missing values

correlation_belsky <- cor.test(correlation_data_belsky$kdm, correlation_data_belsky$poa)

# Calculate correlation and p-value
correlation_data_belsky_poa <- belsky_kdm_poa %>%
  select(poa.x, poa.y) %>%
  drop_na()  # Remove rows with missing values

correlation_belsky_poa <- cor.test(correlation_data_belsky_poa$poa.x, correlation_data_belsky_poa$poa.y

# Calculate correlation and p-value
correlation_data_cb <- chang_belsky %>%
  select(poa.x, poa.y) %>%
  drop_na()  # Remove rows with missing values

correlation_cb <- cor.test(correlation_data_cb$poa.x, correlation_data_cb$poa.y)

# Scatterplot with regression line and correlation info
ggplot(data = chang_kdm, aes(x = kdm, y = poa)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(
    x = Inf, y = Inf,
    label = paste("Correlation =", round(correlation_chang$estimate, 2),
                  "\np-value =", format.pval(correlation_chang$p.value)),
    hjust = 1, vjust = 1, size = 3
  ) +
  xlab("kdm at Week 96") +
  ylab("poa") +
  ggtitle("Scatterplot of kdm vs. poa with Regression Line")
```

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



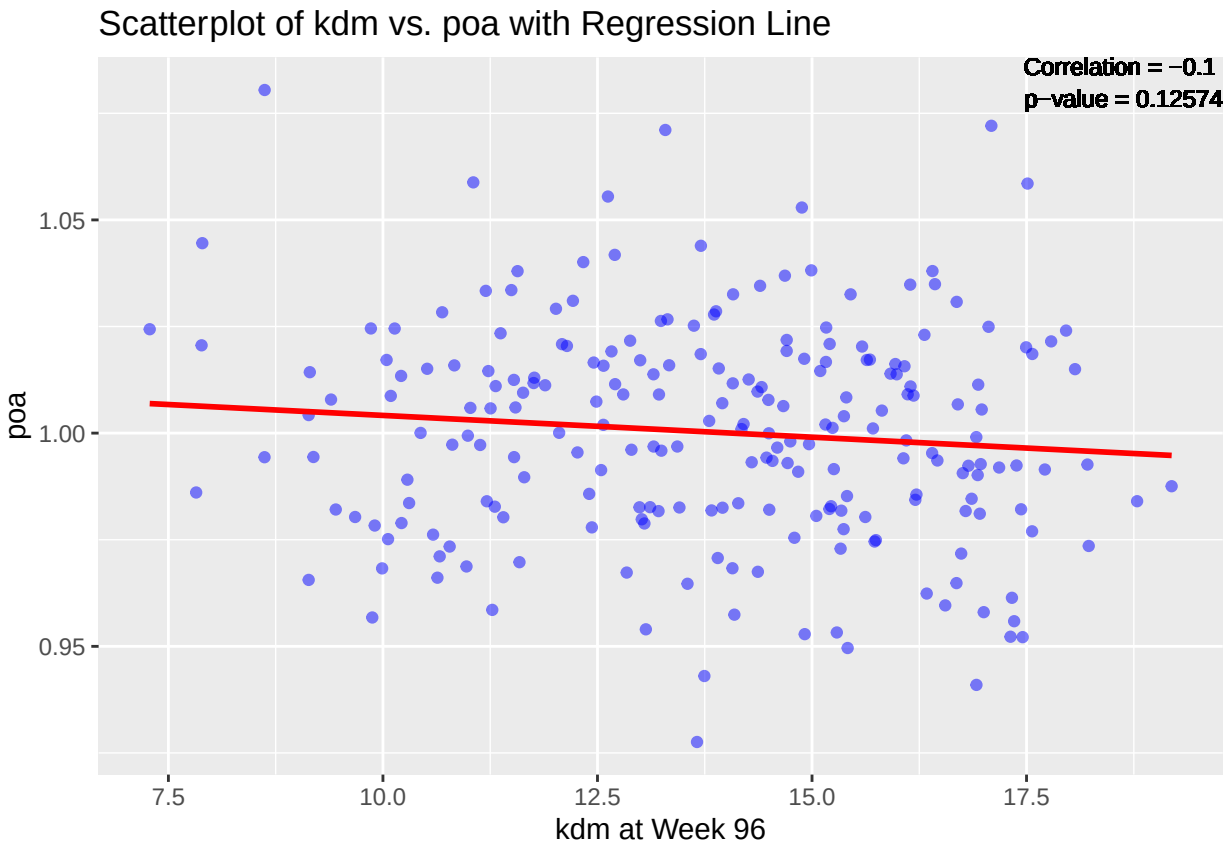

```r
# Scatterplot with regression line and correlation info
ggplot(data = chang_kdm_poa, aes(x = poa.x, y = poa.y)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(
    x = Inf, y = Inf,
    label = paste("Correlation =", round(correlation_chang_poa$estimate, 2),
                  "\np-value =", format.pval(correlation_chang_poa$p.value)),
    hjust = 1, vjust = 1, size = 3
  ) +
  xlab("chang") +
  ylab("kdm_poa") +
  ggtitle("Scatterplot of chang vs. kdm_poa with Regression Line")
```

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



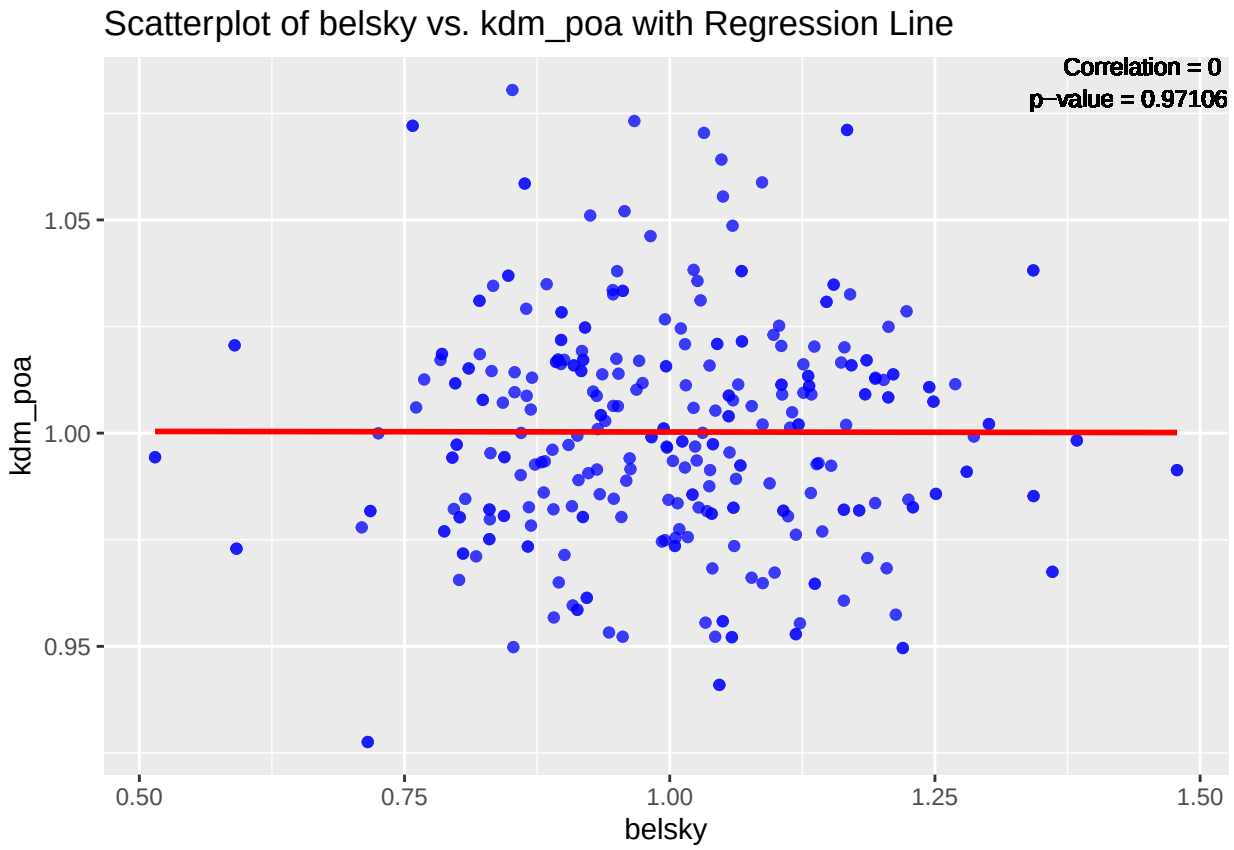

```
# Scatterplot with regression line and correlation info
ggplot(data = belsky_kdm, aes(x = kdm, y = poa)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(
    x = Inf, y = Inf,
    label = paste("Correlation =", round(correlation_belsky$estimate, 2),
                  "\np-value =", format.pval(correlation_belsky$p.value)),
    hjust = 1, vjust = 1, size = 3
  ) +
  xlab("kdm at Week 96") +
  ylab("poa") +
  ggtitle("Scatterplot of kdm vs. poa with Regression Line")
```

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



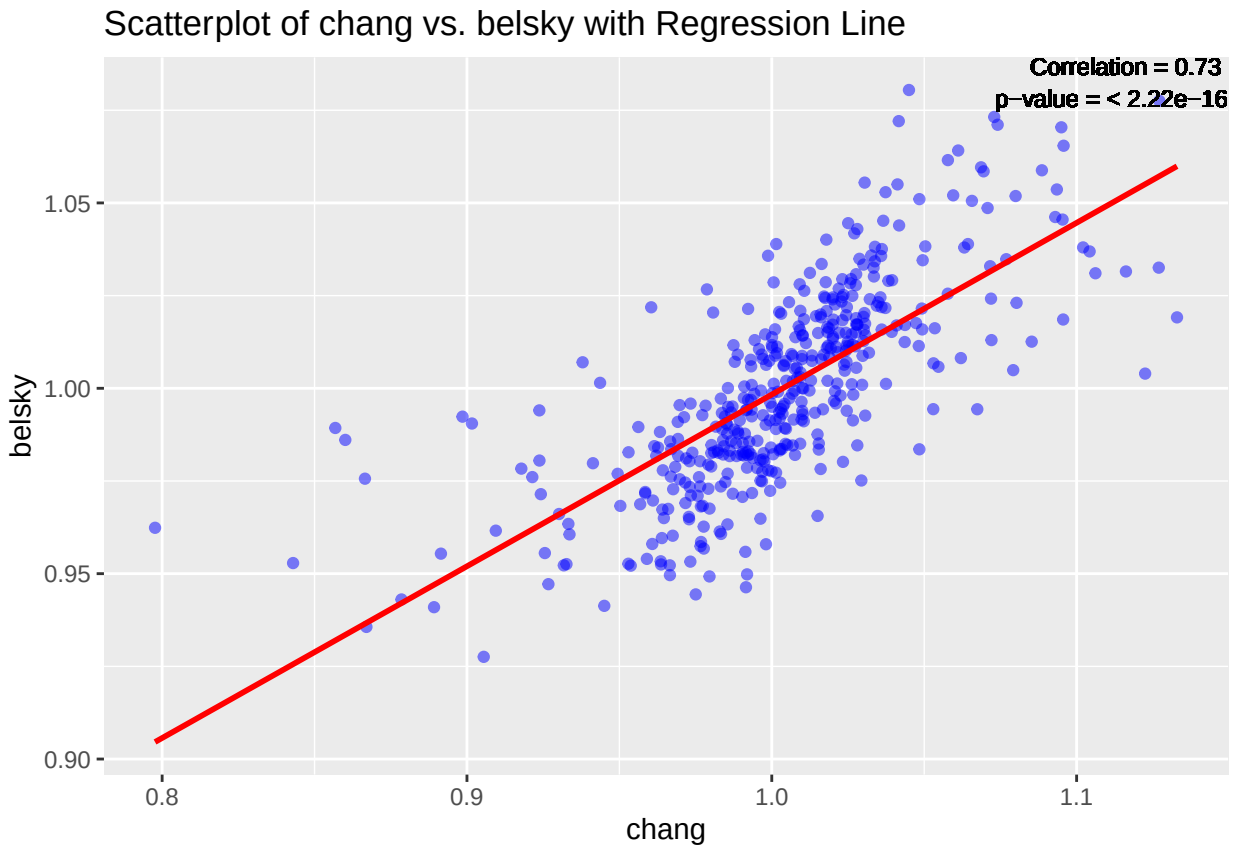

```
# Scatterplot with regression line and correlation info
ggplot(data = belsky_kdm_poa, aes(x = poa.x, y = poa.y)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(
    x = Inf, y = Inf,
    label = paste("Correlation =", round(correlation_belsky_poa$estimate, 2),
                  "\np-value =", format.pval(correlation_belsky_poa$p.value)),
    hjust = 1, vjust = 1, size = 3
  ) +
  xlab("belsky") +
  ylab("kdm_poa") +
  ggtitle("Scatterplot of belsky vs. kdm_poa with Regression Line")
```

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



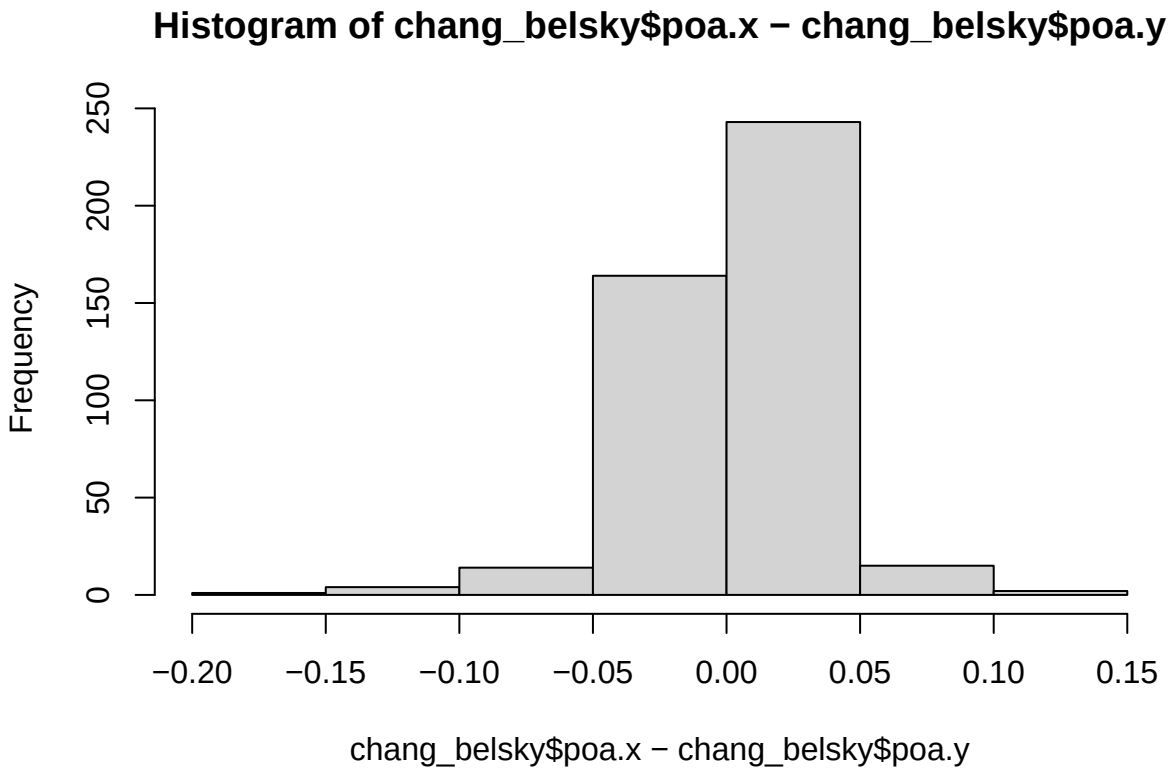

```
# Scatterplot with regression line and correlation info
ggplot(data = chang_belsky, aes(x = poa.x, y = poa.y)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  geom_text(
    x = Inf, y = Inf,
    label = paste("Correlation =", round(correlation_cb$estimate, 2),
                  "\np-value =", format.pval(correlation_cb$p.value)),
    hjust = 1, vjust = 1, size = 3
  ) +
  xlab("chang") +
  ylab("belsky") +
  ggtitle("Scatterplot of chang vs. belsky with Regression Line")
```

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



Differences between PoA

```
hist(chang_belsky$poa.x - chang_belsky$poa.y)
```




```
hist(scale(chang_belsky$poa.x) - scale(chang_belsky$poa.y))
```