

Comparing buffalo and simulated data using summary statistics

Scott Forrest

2024-03-19

Load packages

```
options(scipen=999)

library(tidyverse)
packages <- c("ggh4x", "patchwork", "terra", "ggExtra",
            "cowplot", "ggpubr", "viridis", "scales")
walk(packages, require, character.only = T)
```

Import environmental layers

```
# unscaled rasters
ndvi_stack <- rast("mapping/cropped rasters/ndvi_GEE_projected_watermask20230207.tif")
ndvi_2018_dry <- ndvi_stack[8:10]
ndvi_2019_dry <- ndvi_stack[[19:21]]
ndvi_dry <- terra::mean(c(ndvi_2018_dry, ndvi_2019_dry))
names(ndvi_dry) <- "ndvi_dry"
# plot(ndvi_dry)

canopy <- rast("mapping/cropped rasters/canopy_cover.tif")
herby <- rast("mapping/cropped rasters/veg_herby.tif")
slope <- rast("mapping/cropped rasters/slope_raster.tif")
```

Hourly summary statistics

Import the summary statistic data frames from the observed and simulated datasets that were outputted by the trajectory validation scripts, and combine into a single data frame for plotting

```
# read in the observed data
summaries_hourly_buffalo <-
  read_csv("outputs/buffalo_summaries_hourly_habitat_2024-02-07.csv") %>%
  mutate(id = as.factor(id))

# read in the simulated data
summaries_hourly_0p <-
  read_csv("outputs/sim_0p_memALL_summaries_hourly_habitat_2024-02-07.csv") %>%
  mutate(id = as.factor(id))
summaries_hourly_1p <-
  read_csv("outputs/sim_1p_memALL_summaries_hourly_habitat_2024-02-07.csv") %>%
  mutate(id = as.factor(id))
```

```

summaries_hourly_2p <-
  read_csv("outputs/sim_2p_memALL_summaries_hourly_habitat_2024-02-07.csv") %>%
  mutate(id = as.factor(id))
summaries_hourly_3p <-
  read_csv("outputs/sim_3p_memALL_summaries_hourly_habitat_2024-02-07.csv") %>%
  mutate(id = as.factor(id))

summaries_hourly_all <- bind_rows(summaries_hourly_buffalo,
                                    summaries_hourly_0p,
                                    summaries_hourly_1p,
                                    summaries_hourly_2p,
                                    summaries_hourly_3p)

summaries_hourly_all_long <- summaries_hourly_all %>%
  dplyr::select(!...1) %>%
  pivot_longer(cols = !c(data, hour, id),
               names_to = "variable",
               values_to = "value") %>%
  mutate(Data = factor(str_to_title(data),
                       levels = c("Buffalo", "0p", "1p", "2p", "3p")))

head(summaries_hourly_all)

## # A tibble: 6 x 23
##   ...1 data     hour id      n step_length_mean step_length_median step_length_sd ndvi_mean ndvi_m...
##   <dbl> <chr>    <dbl> <fct> <dbl>           <dbl>           <dbl>           <dbl>       <dbl>
## 1     1 buffalo    1 2005    99      63.7          7.61        117.      0.281
## 2     2 buffalo    1 2014    75     149.          9.82        311.      0.267
## 3     3 buffalo    1 2018    96     206.          15.6        413.      0.263
## 4     4 buffalo    1 2021    97     129.          84.0        214.      0.219
## 5     5 buffalo    1 2022    86     106.          18.8        169.      0.293
## 6     6 buffalo    1 2024    87     265.          71.4        567.      0.297
## # i 9 more variables: canopy_mean <dbl>, canopy_median <dbl>, canopy_sd <dbl>, slope_mean <dbl>, slope...
## #   memory_mean <dbl>, memory_median <dbl>, memory_sd <dbl>

head(summaries_hourly_all_long)

## # A tibble: 6 x 6
##   data     hour id   variable      value Data
##   <chr>    <dbl> <fct> <chr>     <dbl> <fct>
## 1 buffalo    1 2005 n            99    Buffalo
## 2 buffalo    1 2005 step_length_mean 63.7  Buffalo
## 3 buffalo    1 2005 step_length_median 7.61 Buffalo
## 4 buffalo    1 2005 step_length_sd    117. Buffalo
## 5 buffalo    1 2005 ndvi_mean      0.281 Buffalo
## 6 buffalo    1 2005 ndvi_median    0.268 Buffalo

```

Plot the hourly habitat selection between the observed and simulated data from different models

In these plots we are assessing how well the simulated trajectories capture the movement dynamics and habitat use of the observed data.

There's a few different ways we can plot this, so we'll provide some examples here.

With each hour represented as a boxplot

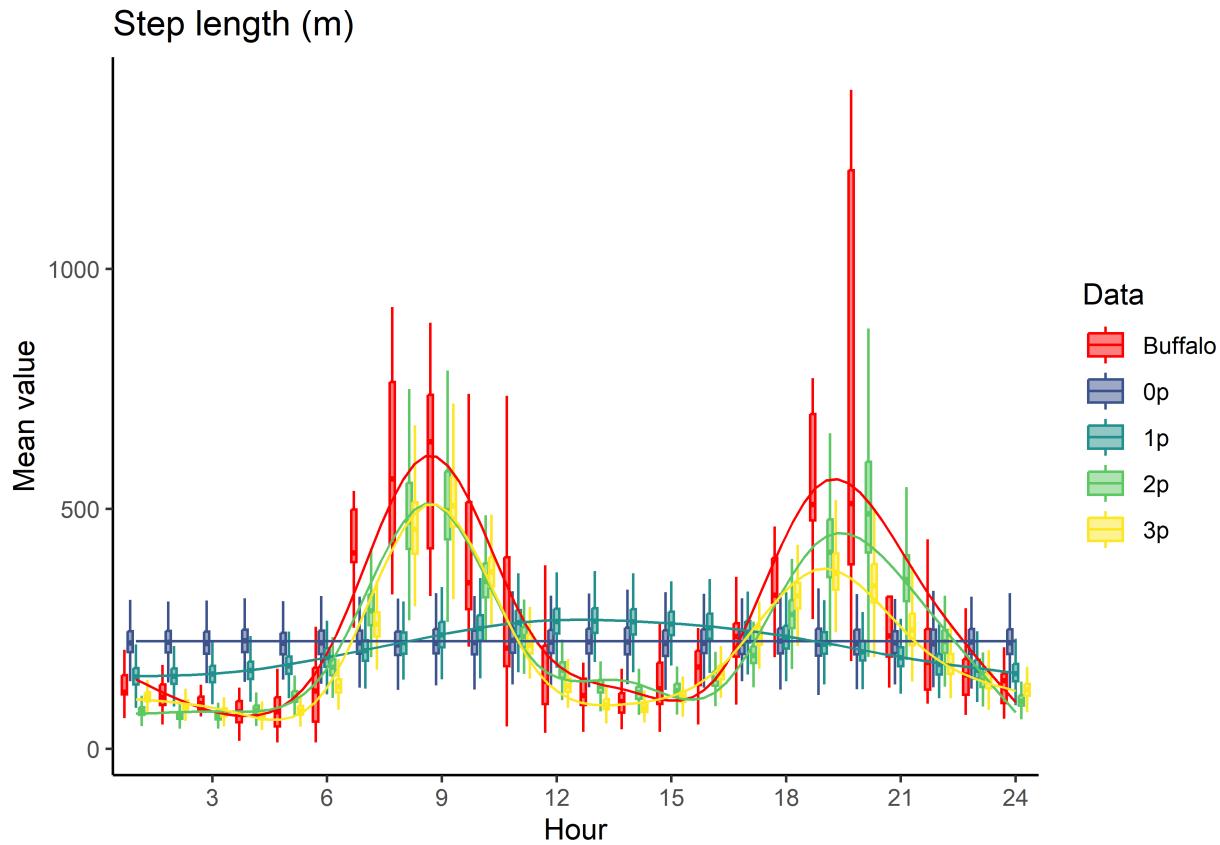
Here, each ‘point’ in a boxplot represents the mean value of step length for that hour for one of the trajectories (observed or simulated). Therefore, the boxplot represents the distribution of these means across all trajectories for each of the models or the observed data. We also include a smoothed line to show the trend.

The boxplots get hard to interpret with the habitat covariates.

```
# Create color mapping
unique_groups <- unique(summaries_hourly_all_long$Data)
colors <- viridis(length(unique_groups))
names(colors) <- unique_groups
colors["Buffalo"] <- "red"

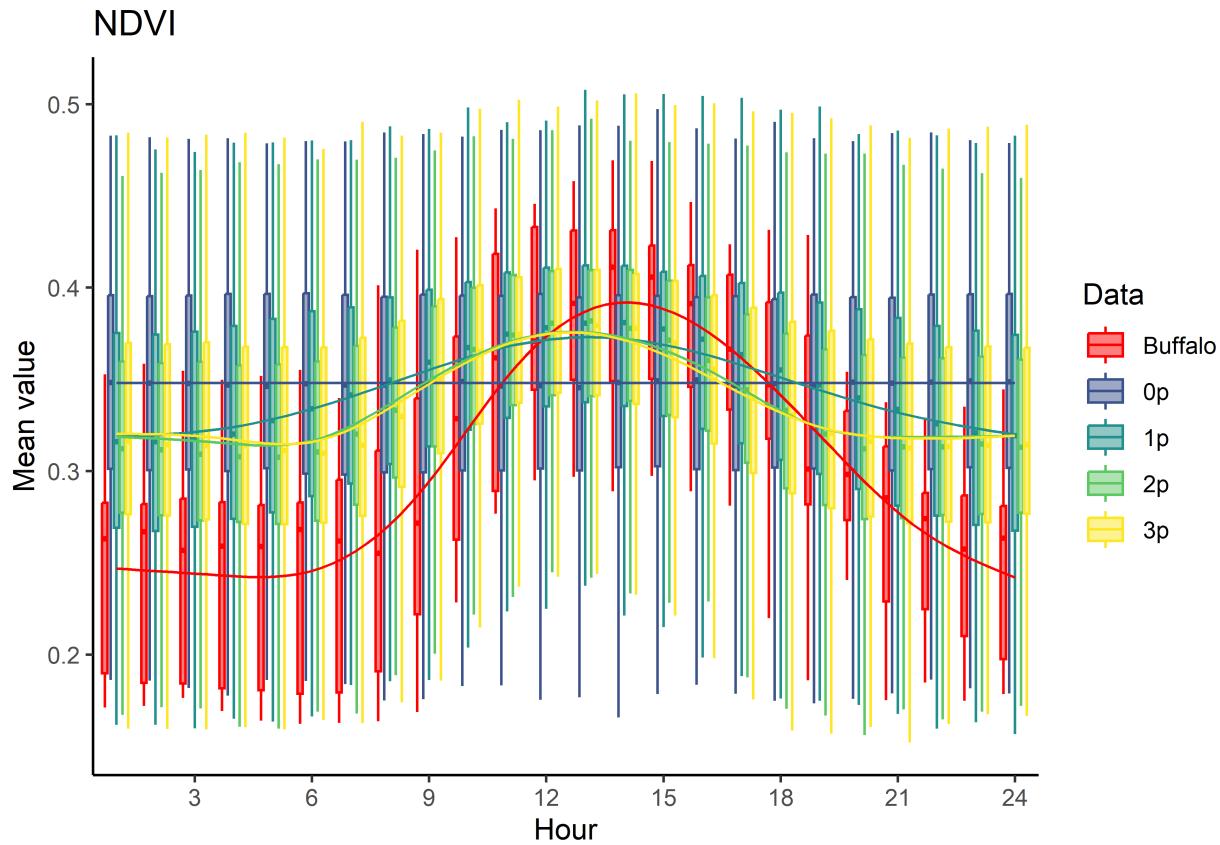
# Step length
ggplot() +
  geom_boxplot(data = summaries_hourly_all_long %>% filter(variable == "step_length_mean"),
               aes(x = factor(hour), y = value, colour = Data, fill = Data),
               alpha = 0.5,
               outlier.shape = NA) +
  geom_smooth(data = summaries_hourly_all_long %>% filter(variable == "step_length_mean"),
              aes(x = hour, y = value, colour = Data, fill = Data),
              size = 0.5,
              se = FALSE,
              alpha = 0.75) +
  scale_fill_manual(values = colors) +
  scale_colour_manual(values = colors) +
  scale_x_discrete("Hour", breaks = seq(0,24,3)) +
  scale_y_continuous("Mean value") +
  ggtitle("Step length (m)") +
  theme_classic()

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



```
# NDVI
ggplot() +
  geom_boxplot(data = summaries_hourly_all_long %>% filter(variable == "ndvi_mean"),
               aes(x = factor(hour), y = value, colour = Data, fill = Data),
               alpha = 0.5,
               outlier.shape = NA) +
  geom_smooth(data = summaries_hourly_all_long %>% filter(variable == "ndvi_mean"),
              aes(x = hour, y = value, colour = Data, fill = Data),
              size = 0.5,
              se = FALSE,
              alpha = 0.75) +
  scale_fill_manual(values = colors) +
  scale_colour_manual(values = colors) +
  scale_x_discrete("Hour", breaks = seq(0,24,3)) +
  scale_y_continuous("Mean value") +
  ggtitle("NDVI") +
  theme_classic()

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



With a path approach

Here we show a single ‘path’ per observed or simulated individual, which is comprised of the mean step length (or other covariate) for each hour. We also include a smoothed line to show the trend. The 13 observed buffalo are shown in red, and the simulated trajectories are coloured based on the model that was fitted (0p, 1p, 2p or 3p to denote 0 - 3 pairs of harmonics).

Again, these are a bit messy, but we can see every trajectory. It’s clear that there is a lot of variability in both the observed and simulated data, which is likely due to where in the landscape the observed or simulated buffalo were (and what habitat is ‘available’ to them). We can see that the dynamic models (2p and 3p for the step length, and 1p, 2p and 3p for NDVI) capture the dynamic trends in movement behaviour and habitat selection.

```
# subset the number of simulations
#(650 per model in our case = 13 buffalo x 50 simulations)
summaries_hourly_all_long_subset <- summaries_hourly_all_long %>%
  dplyr::group_by(data, hour, variable) %>% slice_head(n = 650)

summaries_hourly_all_long_subset %>% dplyr::group_by(data, id)

## # A tibble: 1,191,528 x 6
## # Groups:   data, id [2,613]
##   data    hour   id     variable     value Data
##   <chr>  <dbl> <fct>    <chr>      <dbl> <fct>
## 1 0p        1 2005_10_1 canopy_mean 0.570 0p
## 2 0p        1 2005_10_2 canopy_mean 0.537 0p
## 3 0p        1 2005_11_2 canopy_mean 0.444 0p
```

```

## 4 0p      1 2005_12_2 canopy_mean 0.514 0p
## 5 0p      1 2005_13_2 canopy_mean 0.494 0p
## 6 0p      1 2005_14_2 canopy_mean 0.334 0p
## 7 0p      1 2005_15_2 canopy_mean 0.527 0p
## 8 0p      1 2005_16_2 canopy_mean 0.376 0p
## 9 0p      1 2005_17_2 canopy_mean 0.519 0p
## 10 0p     1 2005_18_2 canopy_mean 0.564 0p
## # i 1,191,518 more rows

# set path alpha
sim_path_alpha <- 0.1
buff_path_alpha <- 0.25

# set smooth alpha
sim_smooth_alpha <- 1
buff_smooth_alpha <- 1

# linewidth
sim_path_linewidth <- 0.25
buff_path_linewidth <- 0.5

# smooth linewidth
sim_smooth_linewidth <- 1
buff_smooth_linewidth <- 1

# Step length
ggplot() +

  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(!Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = sim_path_alpha,
    linewidth = sim_path_linewidth) +
  
  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +
  
  geom_smooth(data = summaries_hourly_all_long %>%
    filter(!Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = value, colour = Data, fill = Data),
    method = "gam",
    size = sim_smooth_linewidth,
    se = FALSE,
    alpha = sim_smooth_alpha) +
  
  geom_smooth(data = summaries_hourly_all_long %>%
    filter(Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = value, colour = Data, fill = Data),
    method = "gam",
    size = buff_smooth_linewidth,
    se = FALSE,
    alpha = buff_smooth_alpha) +

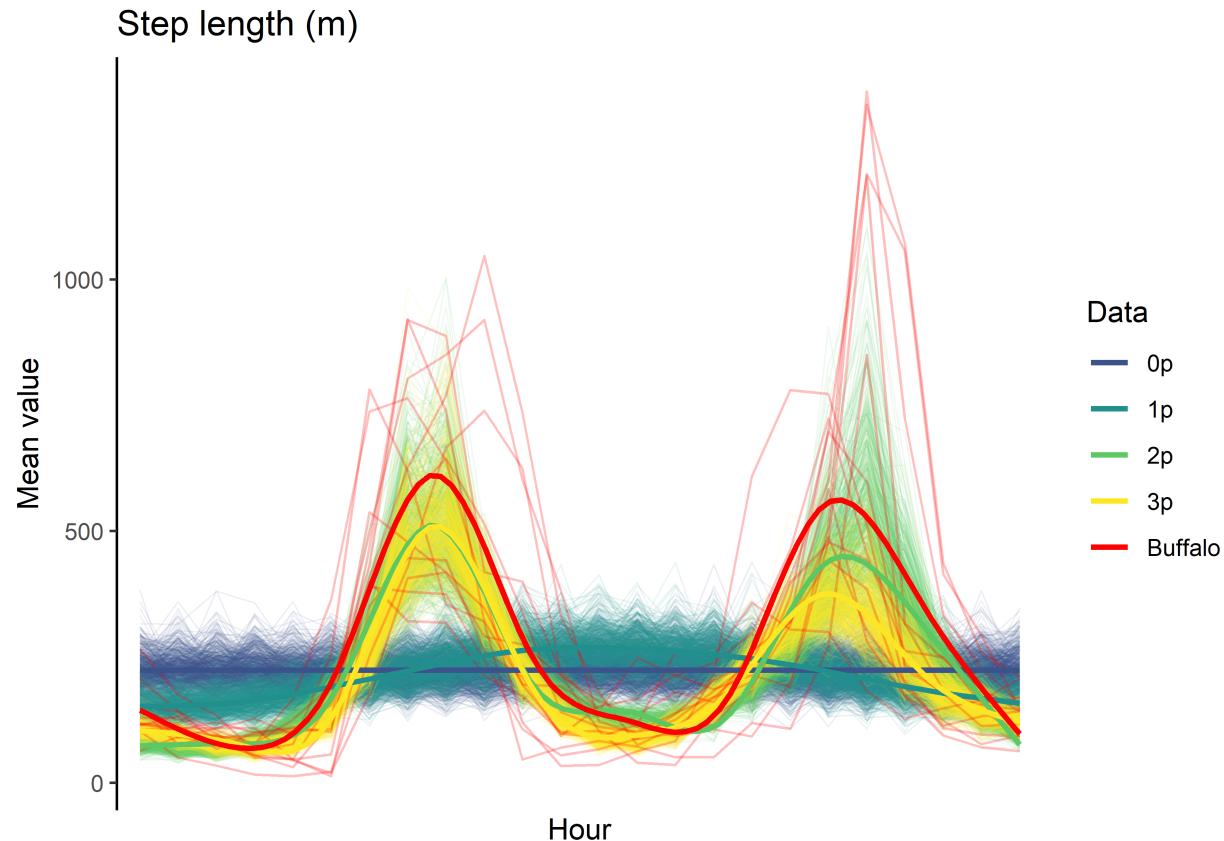
```

```

scale_fill_manual(values = colors) +
scale_colour_manual(values = colors) +
scale_x_discrete("Hour", breaks = seq(0,24,3)) +
scale_y_continuous("Mean value") +
ggtitle("Step length (m)") +
theme_classic()

## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'

```



```

# NDVI
ggplot() +

  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(!Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = sim_path_alpha,
    linewidth = sim_path_linewidth) +

  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +

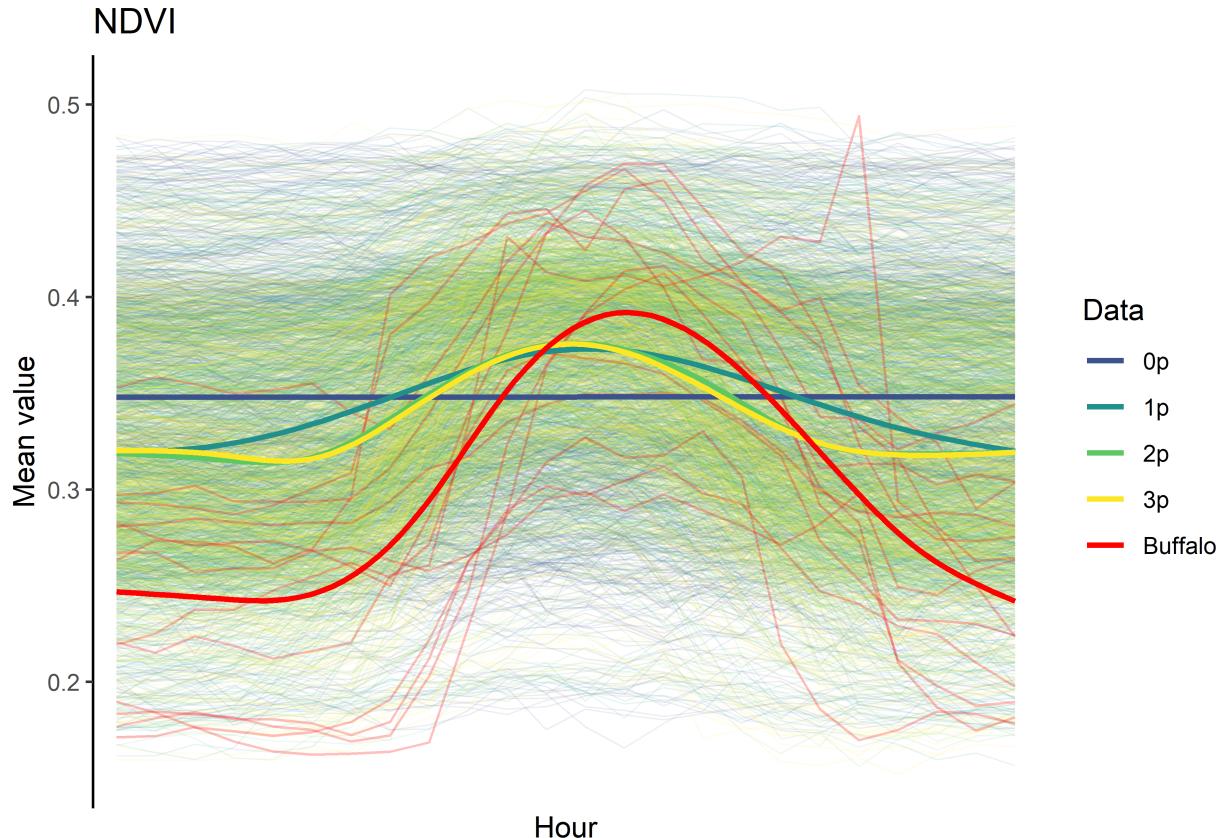
```

```

geom_smooth(data = summaries_hourly_all_long %>%
  filter(!Data == "Buffalo" & variable == "ndvi_mean"),
  aes(x = hour, y = value, colour = Data, fill = Data),
  method = "gam",
  size = sim_smooth_linewidth,
  se = FALSE,
  alpha = sim_smooth_alpha) +
geom_smooth(data = summaries_hourly_all_long %>%
  filter(Data == "Buffalo" & variable == "ndvi_mean"),
  aes(x = hour, y = value, colour = Data, fill = Data),
  method = "gam",
  size = buff_smooth_linewidth,
  se = FALSE,
  alpha = buff_smooth_alpha) +
scale_fill_manual(values = colors) +
scale_colour_manual(values = colors) +
scale_x_discrete("Hour", breaks = seq(0,24,3)) +
scale_y_continuous("Mean value") +
ggtitle("NDVI") +
theme_classic()

## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using formula = 'y ~ s(x, bs = "cs")'

```



A quantile ribbon approach

To express the stochasticity of the simulations, here we show the 25th to 50th quantiles and the 2.5th to 97.5th quantiles of the data. Remember that the ‘data’ are the means for each hour for each trajectory, so the quantiles are calculated across the means for each hour.

This is the plotting approach that we used in the paper.

Here we create a ribbon from the 25th to 50th quantiles and from the 2.5th to 97.5th quantiles of the data

```
hourly_summary_quantiles <- summaries_hourly_all_long %>%
  dplyr::group_by(Data, hour, variable) %>%
  summarise(n = n(),
            mean = mean(value, na.rm = TRUE),
            sd = sd(value, na.rm = TRUE),
            q025 = quantile(value, probs = 0.025, na.rm = TRUE),
            q25 = quantile(value, probs = 0.25, na.rm = TRUE),
            q50 = quantile(value, probs = 0.5, na.rm = TRUE),
            q75 = quantile(value, probs = 0.75, na.rm = TRUE),
            q975 = quantile(value, probs = 0.975, na.rm = TRUE))
```

‘summarise()’ has grouped output by ‘Data’, ‘hour’. You can override using the ‘.groups’ argument.

Here we show the hourly step length and selection of three of the habitat covariates. We use a dashed-line ribbon for the 95% interval and a solid-line ribbon for the 50% interval. We show the mean as a solid line.

It’s quite a long plotting script...

```
# set plotting parameters here that will change in each plot
buff_path_alpha <- 0.1
ribbon_95_alpha <- 0.1
ribbon_50_alpha <- 0.15
path_95_alpha <- 1

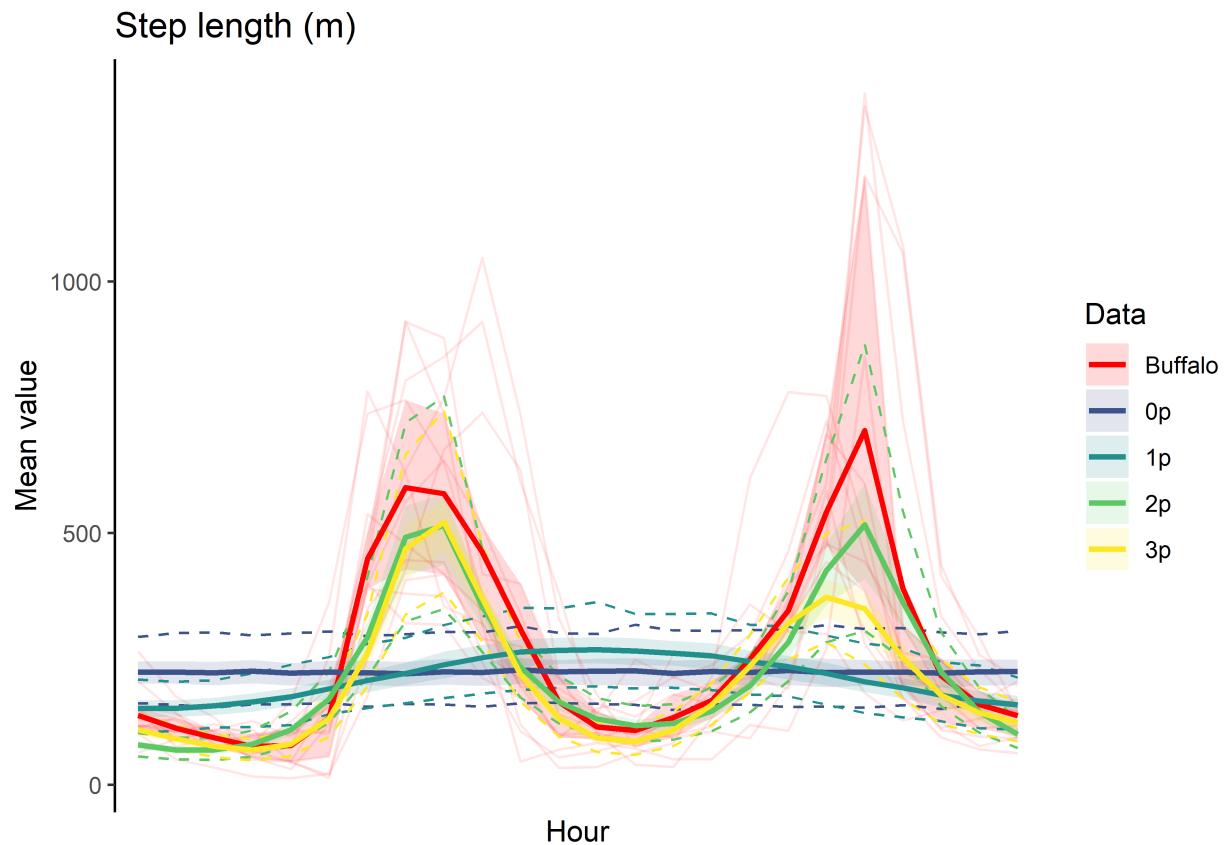
hourly_path_sl_plot <- ggplot() +
  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +
  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +
  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +
  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = q025, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +
```

```

geom_path(data = hourly_summary_quantiles %>%
  filter(!Data == "Buffalo" & variable == "step_length_mean"),
  aes(x = hour, y = q975, colour = Data),
  linetype = "dashed",
  alpha = path_95_alpha) +
  geom_path(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +
  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "step_length_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +
  scale_fill_manual(values = colors) +
  scale_colour_manual(values = colors) +
  scale_x_discrete("Hour", breaks = seq(0,24,3)) +
  scale_y_continuous("Mean value") +
  ggtitle("Step length (m)") +
  theme_classic()

hourly_path_sl_plot

```



```

hourly_path_ndvi_plot <- ggplot() +

  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +

  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +

  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +

  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = q025, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +

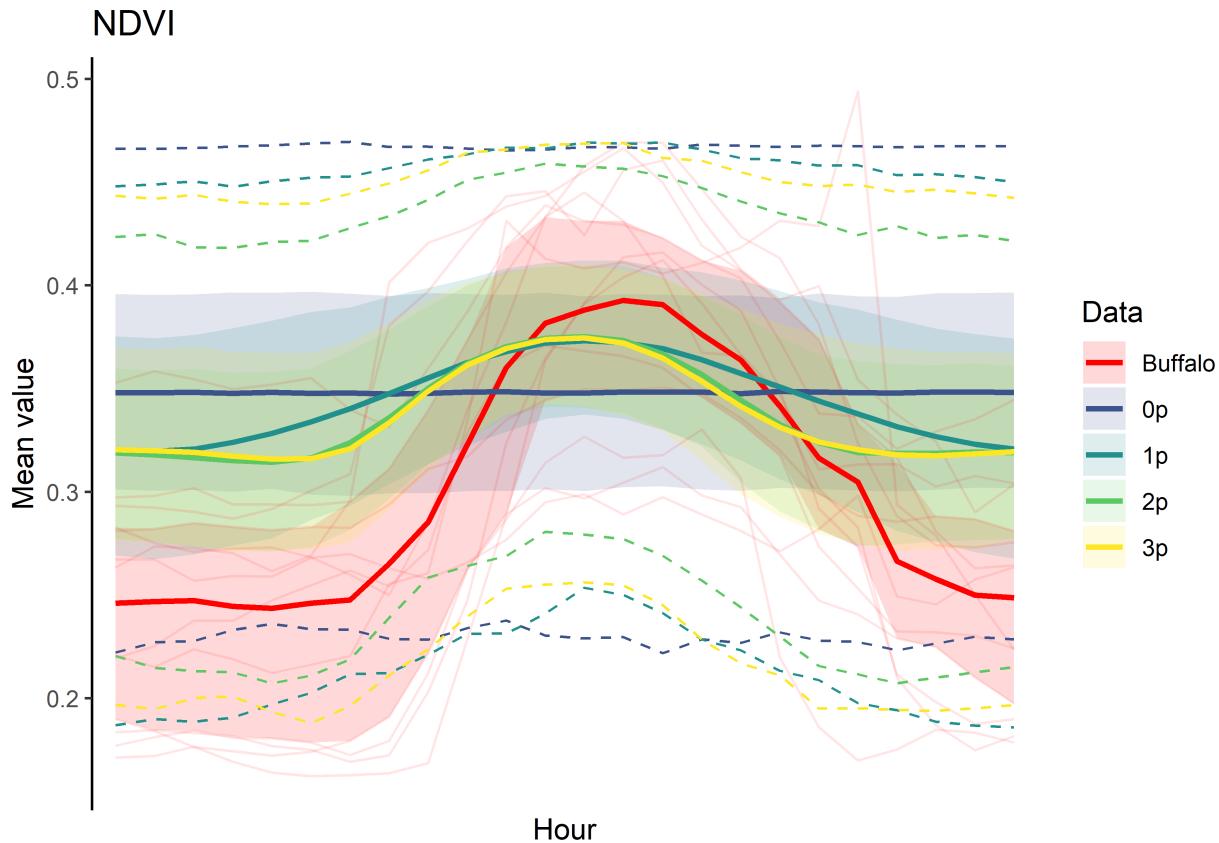
  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = q975, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +

  geom_path(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +

  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "ndvi_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +
  scale_fill_manual(values = colors) +
  scale_colour_manual(values = colors) +
  scale_x_discrete("Hour", breaks = seq(0,24,3)) +
  scale_y_continuous("Mean value") +
  ggtitle("NDVI") +
  theme_classic()

hourly_path_ndvi_plot

```



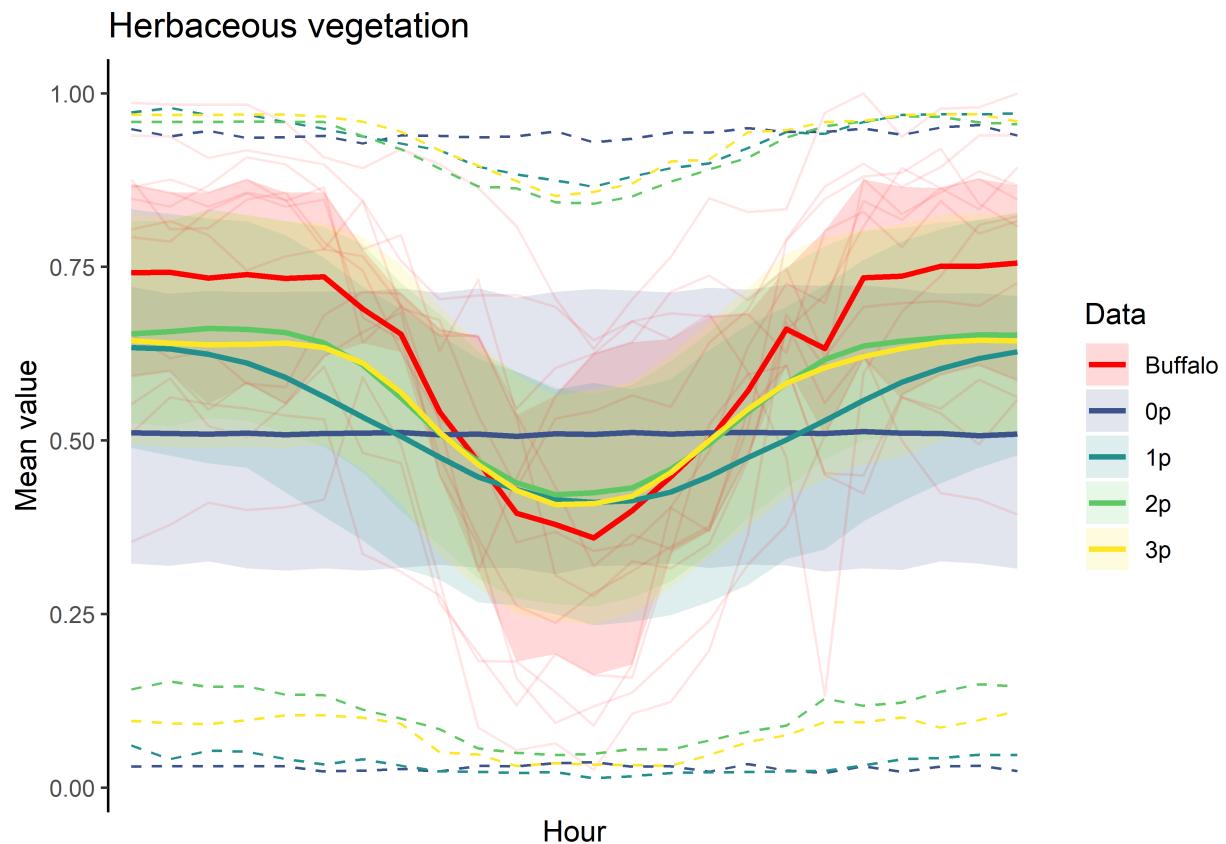
```
hourly_path_herby_plot <- ggplot() +
  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "herby_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +
  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "herby_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +
  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "herby_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +
  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "herby_mean"),
    aes(x = hour, y = q025, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +
  geom_path(data = hourly_summary_quantiles %>%
```

```

filter(!Data == "Buffalo" & variable == "herby_mean"),
aes(x = hour, y = q975, colour = Data),
linetype = "dashed",
alpha = path_95_alpha) +
geom_path(data = hourly_summary_quantiles %>%
  filter(Data == "Buffalo" & variable == "herby_mean"),
  aes(x = hour, y = mean, colour = Data),
  linewidth = 1) +
geom_path(data = hourly_summary_quantiles %>%
  filter(!Data == "Buffalo" & variable == "herby_mean"),
  aes(x = hour, y = mean, colour = Data),
  linewidth = 1) +
scale_fill_manual(values = colors) +
scale_colour_manual(values = colors) +
scale_x_discrete("Hour", breaks = seq(0,24,3)) +
scale_y_continuous("Mean value") +
ggtitle("Herbaceous vegetation") +
theme_classic()

hourly_path_herby_plot

```



```

hourly_path_canopy_plot <- ggplot() +  

  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +  

  geom_ribbon(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, ymin = q25, ymax = q75, fill = Data),
    alpha = ribbon_50_alpha) +  

  geom_path(data = summaries_hourly_all_long_subset %>%
    filter(Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, y = value, colour = Data, group = interaction(id, Data)),
    alpha = buff_path_alpha,
    linewidth = buff_path_linewidth) +  

  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, y = q025, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +  

  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, y = q975, colour = Data),
    linetype = "dashed",
    alpha = path_95_alpha) +  

  geom_path(data = hourly_summary_quantiles %>%
    filter(Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +  

  geom_path(data = hourly_summary_quantiles %>%
    filter(!Data == "Buffalo" & variable == "canopy_mean"),
    aes(x = hour, y = mean, colour = Data),
    linewidth = 1) +  

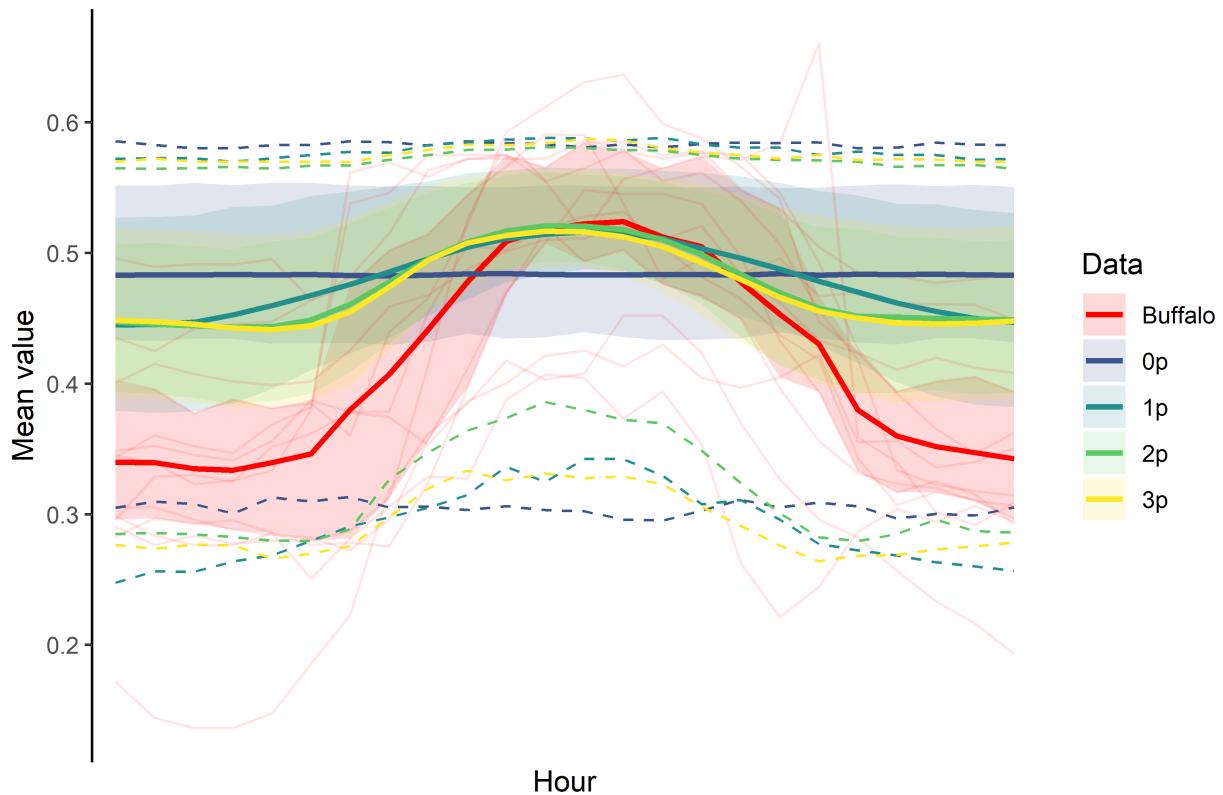
  scale_fill_manual(values = colors) +
  scale_colour_manual(values = colors) +
  scale_x_discrete("Hour", breaks = seq(0,24,3)) +
  scale_y_continuous("Mean value") +
  ggtitle("Canopy cover") +
  theme_classic()  

hourly_path_canopy_plot

```

Canopy cover



Combining the hourly plots

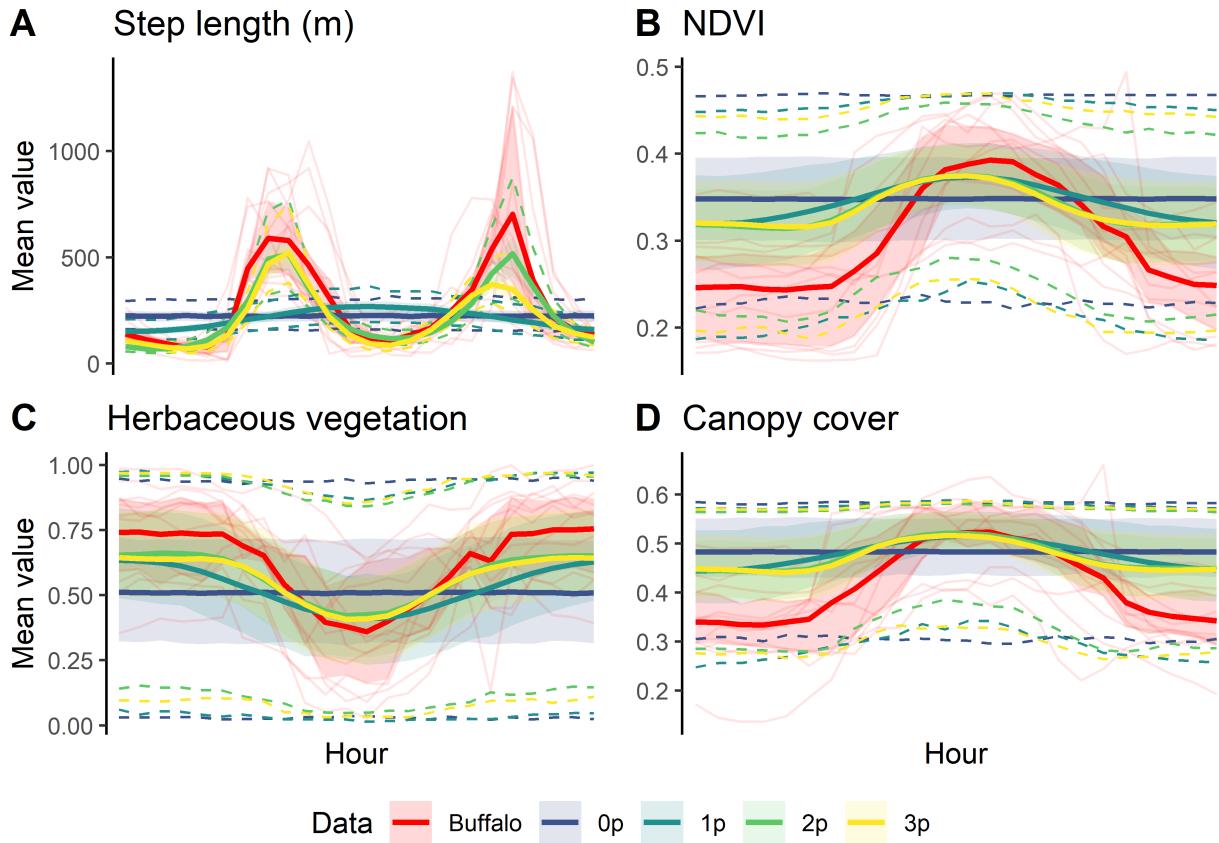
```
ggarrange(hourly_path_sl_plot +
  theme(axis.title.x = element_blank(),
        axis.text.x = element_blank()),

  hourly_path_ndvi_plot +
  theme(axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        axis.title.y = element_blank()),

  hourly_path_herby_plot,

  hourly_path_canopy_plot +
  theme(axis.title.y = element_blank()),

  labels = c("A", "B", "C", "D"),
  ncol = 2, nrow = 2,
  legend = "bottom",
  common.legend = TRUE)
```



```
ggsave(paste0("outputs/plots/manuscript_figs/memALL_hourly_path_mean_all_",
              Sys.Date(), ".png"),
       width=150, height=120, units="mm", dpi = 600)
```

Full trajectory summary statistics

Here we compare between the observed and simulated trajectories using summary statistics that result in a single value per trajectory (rather than one per hour as above). These summary statistics were also used by Abrahms et al. (2017) and Kays et al. (2023), although to compare between taxa rather than between observed and simulated data.

Import the summary statistic data frames from the observed and simulated datasets and combine into a single data frame for plotting

```
summaries_buffalo <-
  read_csv("outputs/buffalo_summary_statistics_df_2024-02-07.csv")

summaries_0p <-
  read_csv("outputs/sim_0p_memALL_daily_summary_statistics_df_2024-02-07.csv")
summaries_0p <- summaries_0p %>% mutate(sim = "0p")

summaries_1p <-
  read_csv("outputs/sim_1p_memALL_daily_summary_statistics_df_2024-02-07.csv")
summaries_1p <- summaries_1p %>% mutate(sim = "1p")
```

```

summaries_2p <-
  read_csv("outputs/sim_2p_memALL_daily_summary_statistics_df_2024-02-07.csv")
summaries_2p <- summaries_2p %>% mutate(sim = "2p")

summaries_3p <-
  read_csv("outputs/sim_3p_memALL_daily_summary_statistics_df_2024-02-07.csv")
summaries_3p <- summaries_3p %>% mutate(sim = "3p")

summaries_all <- bind_rows(summaries_buffalo,
                            summaries_0p,
                            summaries_1p,
                            summaries_2p,
                            summaries_3p)

summaries_all <- summaries_all %>% mutate(residence_time_log = log(residence_time),
                                             msd_log = log(msd))

summaries_all_long <- summaries_all %>%
  pivot_longer(cols = !c(traj, id, data, sim),
               names_to = "summary", values_to = "value") %>%

  mutate(sim = factor(sim, levels = c("obs", "0p", "1p", "2p", "3p"),
                      labels = c("Buffalo", "0p", "1p", "2p", "3p")),

  # categorise the summary statistics for plotting
  category = case_when(summary == "step_length_median" ~ "step",
                        summary == "step_length_mean" ~ "step",
                        summary == "gamma_shape" ~ "step",
                        summary == "gamma_scale" ~ "step",
                        summary == "vm_kappa" ~ "step",

                        summary == "straightness" ~ "path",
                        summary == "sinuosity" ~ "path",

                        summary == "tac" ~ "life_history",
                        summary == "cum_dist" ~ "life_history",
                        summary == "tot_dist" ~ "life_history",
                        summary == "msd" ~ "life_history",
                        summary == "msd_log" ~ "life_history",
                        summary == "intensity_use" ~ "life_history",
                        summary == "residence_time" ~ "life_history",
                        summary == "residence_time_log" ~ "life_history",
                        summary == "hr_area_50" ~ "life_history",
                        summary == "hr_area_75" ~ "life_history",
                        summary == "hr_area_95" ~ "life_history",
                        summary == "hr_monthly_overlap_ba" ~ "life_history",
                        summary == "hr_monthly_overlap_vi" ~ "life_history",
                        summary == "ba_overlap" ~ "life_history",

                        summary == "ndvi_mean" ~ "habitat",
                        summary == "ndvi_median" ~ "habitat",
                        summary == "ndvi_sd" ~ "habitat",
                        summary == "herby_mean" ~ "habitat",

```

```

        summary == "herby_sd" ~ "habitat",
        summary == "canopy_mean" ~ "habitat",
        summary == "canopy_sd" ~ "habitat",
        summary == "slope_mean" ~ "habitat",
        summary == "slope_median" ~ "habitat",
        summary == "slope_sd" ~ "habitat"

    )))

```

`head(summaries_all_long)`

```

## # A tibble: 6 x 7
##   traj      id data sim     summary      value category
##   <chr>     <dbl> <chr> <fct>   <chr>       <dbl> <chr>
## 1 obs      2005 obs Buffalo step_length_median 92.5    step
## 2 obs      2005 obs Buffalo step_length_mean   194.    step
## 3 obs      2005 obs Buffalo ndvi_mean        0.307   habitat
## 4 obs      2005 obs Buffalo ndvi_median      0.294   habitat
## 5 obs      2005 obs Buffalo ndvi_sd         0.0669  habitat
## 6 obs      2005 obs Buffalo herby_mean       0.764   habitat

```

`tail(summaries_all_long)`

```

## # A tibble: 6 x 7
##   traj      id data sim     summary      value category
##   <chr>     <dbl> <chr> <fct>   <chr>       <dbl> <chr>
## 1 2393_9_26 2393 sim  3p    hr_area_95    33.4  life_history
## 2 2393_9_26 2393 sim  3p    hr_monthly_overlap_ba 0.482 life_history
## 3 2393_9_26 2393 sim  3p    hr_monthly_overlap_vi 0.243 life_history
## 4 2393_9_26 2393 sim  3p    ba_overlap        0.290 life_history
## 5 2393_9_26 2393 sim  3p    residence_time_log 1.94   life_history
## 6 2393_9_26 2393 sim  3p    msd_log          15.8   life_history

```

Calculate the mean of the summary statistic values for plotting

```

summary_mean_values <- summaries_all_long %>%
  group_by(summary, sim) %>%
  summarise(mean_value = mean(value, na.rm = TRUE),
            median_value = median(value, na.rm = TRUE))

```

`## ‘summarise()’ has grouped output by ‘summary’. You can override using the ‘.groups’ argument.`

```

summaries_all_long %>%
  group_by(summary) %>%
  summarise(mean = mean(value, na.rm = TRUE),
            sd = sd(value, na.rm = TRUE),
            median = median(value, na.rm = TRUE),
            min = min(value, na.rm = TRUE),
            max = max(value, na.rm = TRUE)) %>%
  arrange(summary)

```

```

## # A tibble: 31 x 6
##   summary      mean       sd   median      min      max
##   <chr>     <dbl>     <dbl>   <dbl>   <dbl>   <dbl>
## 1 ba_overlap  0.267    0.197   0.259  1.43e-9  0.850
## 2 canopy_mean 0.475    0.0743  0.485  2.23e-1  0.610

```

```

## 3 canopy_sd      0.125      0.0432      0.126 1.11e-2      0.278
## 4 cum_dist     477113.    53218.    475607.    3.11e+5 1001533.
## 5 gamma_scale   467.      53.7       461.    3.45e+2    1017.
## 6 gamma_shape    0.461     0.0172     0.460 3.57e-1      0.524
## 7 herby_mean     0.546     0.233      0.574 3.42e-3      0.995
## 8 herby_sd       0.432     0.0833     0.464 5.84e-2      0.500
## 9 hr_area_50     11.9      20.9       4.58 8.22e-1     298.
## 10 hr_area_75    28.5      47.9       10.7  1.73e+0     580.
## # i 21 more rows

```

Step measurements

```

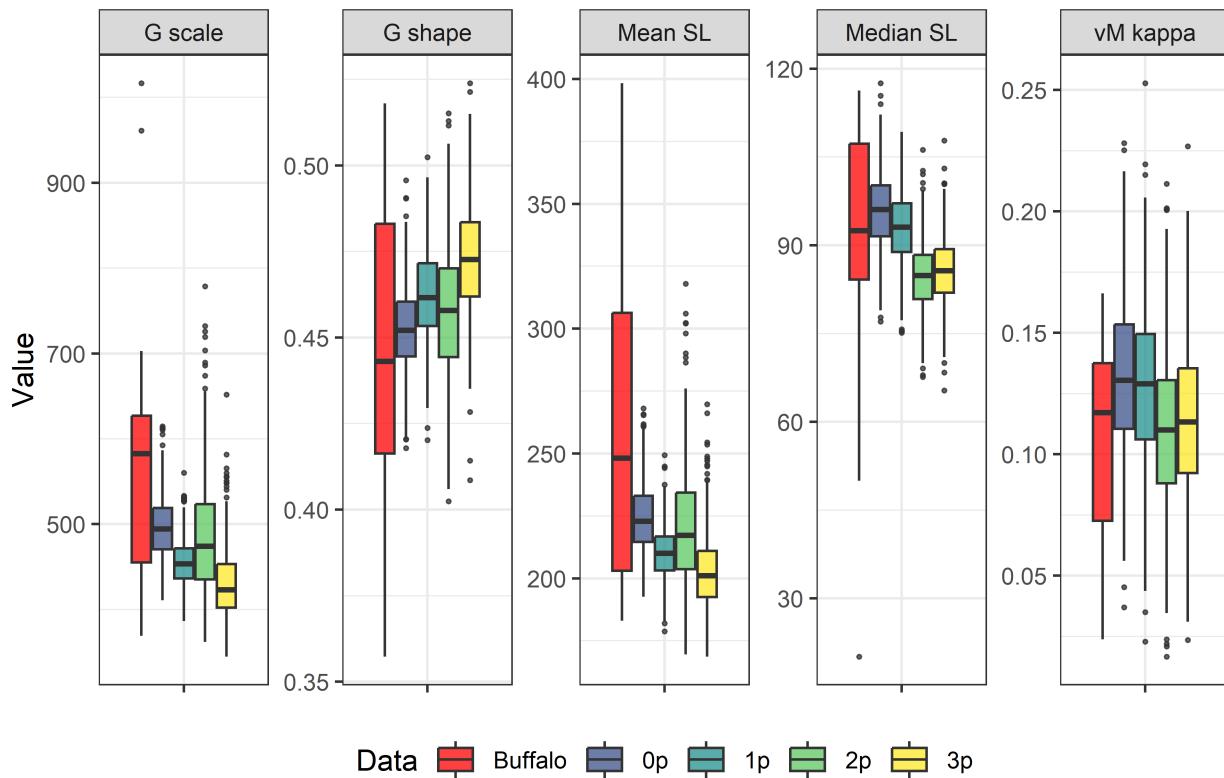
# Create color mapping
unique_groups <- unique(summaries_all_long$sim)
colors <- viridis(length(unique_groups))
names(colors) <- unique_groups
colors["Buffalo"] <- "red"

step_summary_labs <- c("gamma_shape" = "G shape",
                       "gamma_scale" = "G scale",
                       "step_length_mean" = "Mean SL",
                       "step_length_median" = "Median SL",
                       "vm_kappa" = "vM kappa")

ggplot() +
  geom_boxplot(data = summaries_all_long %>%
                filter(category == "step" #&
                      # !str_detect(summary, "gamma")
                      ),
                aes(x = summary, y = value, fill = sim), alpha = 0.75,
                outlier.size = 0.5) +
  ggh4x::facet_grid2(scales = "free", cols = vars(summary),
                     independent = "y",
                     labeller = labeller(summary = step_summary_labs)) +
  ggtitle("Step summary statistics") +
  scale_fill_manual("Data", values = colors) +
  scale_y_continuous("Value") +
  theme_bw() +
  theme(legend.position = "bottom",
        axis.text.x = element_blank(),
        axis.title.x = element_blank())

```

Step summary statistics



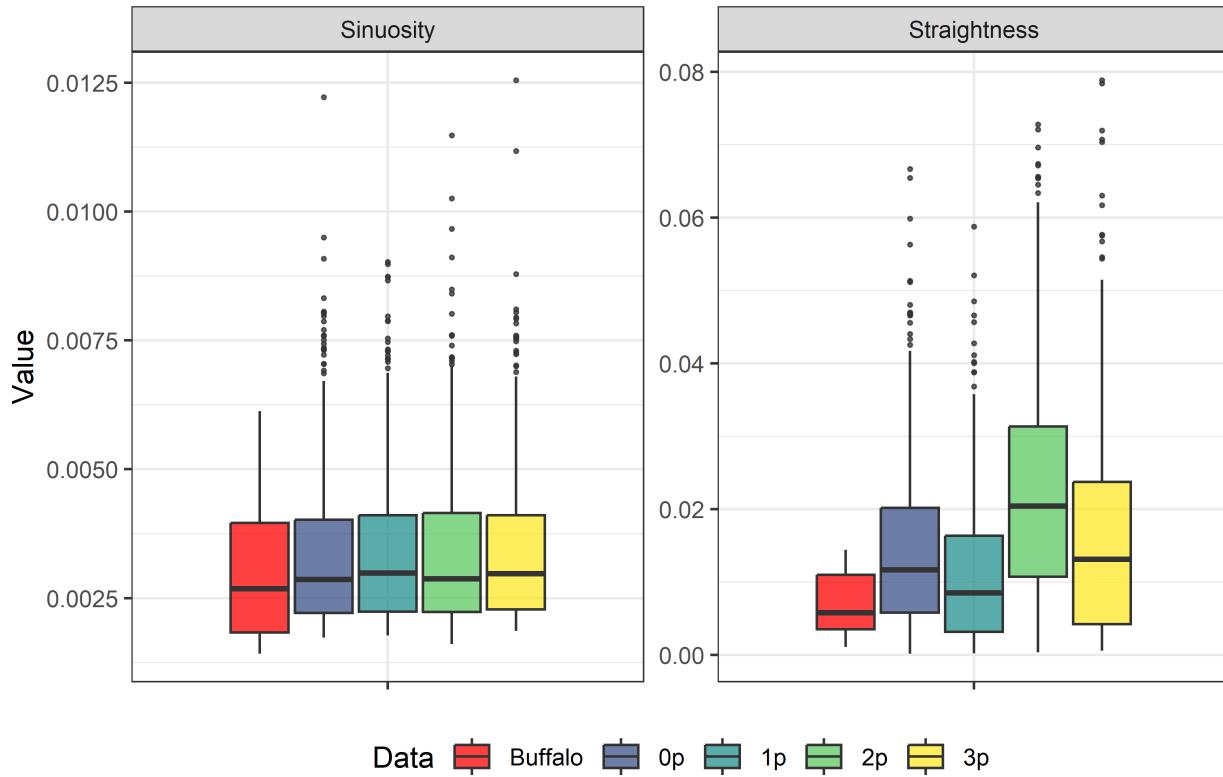
```
# save the plot using ggplot
ggsave(paste0("outputs/plots/manuscript_figs/memALL_summary_comparison_steps_",
              Sys.Date(), ".png"),
       width=150, height=90, units="mm", dpi = 600)
```

Path measurements

```
summary_labs <- c("sinuosity" = "Sinuosity", "straightness" = "Straightness")

ggplot() +
  geom_boxplot(data = summaries_all_long %>%
                filter(category == "path"),
                aes(x = summary, y = value, fill = sim), alpha = 0.75,
                outlier.size = 0.5) +
  ggh4x::facet_grid2(scales = "free", cols = vars(summary), independent = "y"
                     , labeller = labeller(summary = summary_labs)
                     ) +
  ggtitle("Path summary statistics") +
  scale_fill_manual("Data", values = colors) +
  scale_y_continuous("Value") +
  theme_bw() +
  theme(legend.position = "bottom",
        axis.text.x = element_blank(),
        axis.title.x = element_blank())
```

Path summary statistics



```
# save the plot using ggplot
ggsave(paste0("outputs/plots/manuscript_figs/memALL_summary_comparison_paths_",
              Sys.Date(), ".png"),
       width=150, height=90, units="mm", dpi = 600)
```

Life-history measurements

```
lh_summary_labs <- c("intensity_use" = "Intensity of Use",
                      "msd" = "MSD (m^2)",
                      "residence_time" = "RT (hours)",
                      "tac" = "Mean TAC",
                      "tot_dist" = "Tot Dist (m)")

ggplot() +
  geom_boxplot(data = summaries_all_long %>%
                filter(category == "life_history" &
                           (str_detect(summary, "tac") |
                           str_detect(summary, "residence_time") |
                           str_detect(summary, "tot_d") |
                           str_detect(summary, "intensity_use")) &
                           !str_detect(summary, "log")),
                aes(x = summary, y = value, fill = sim), alpha = 0.75,
                outlier.size = 0.5) +
  ggh4x::facet_grid2(scales = "free", cols = vars(summary), independent = "y",
```

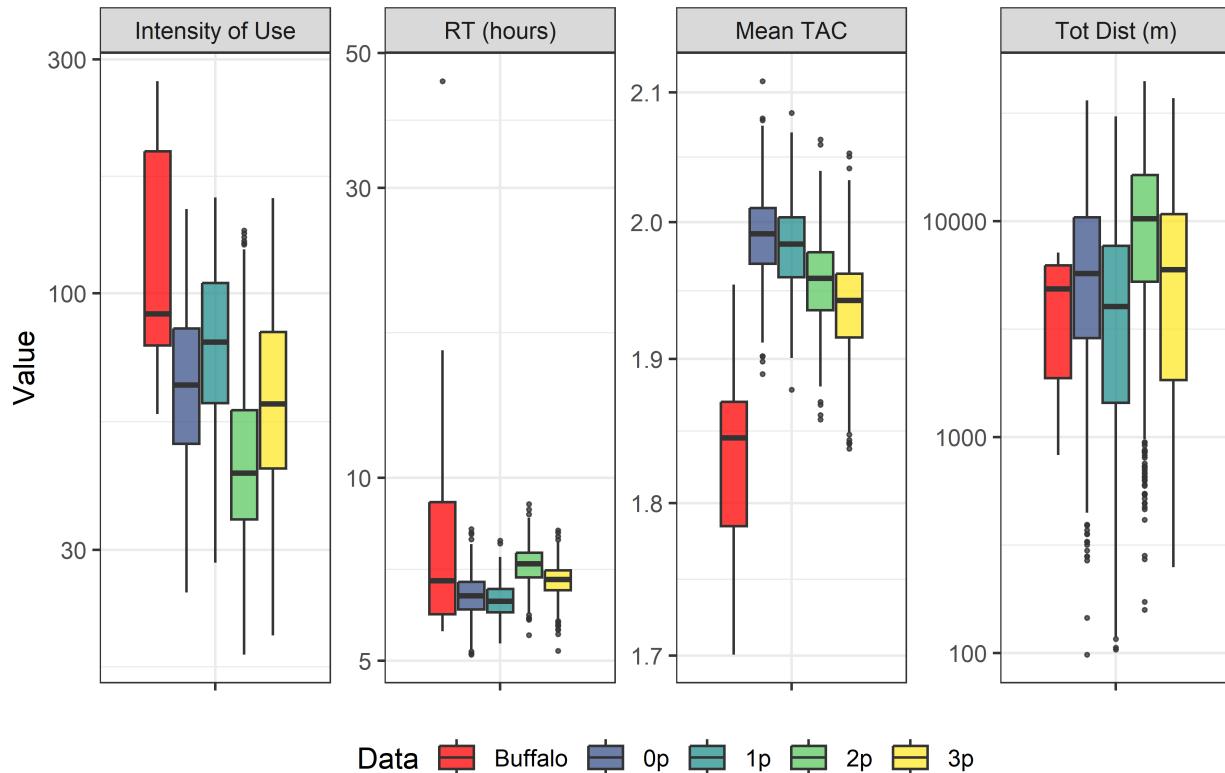
```

    labeller = labeller(summary = lh_summary_labs)) +
ggtitle("Life-history summary statistics") +
scale_fill_manual("Data", values = colors) +
scale_y_log10("Value") +
# scale_y_continuous("Value") +
theme_bw() +
theme(legend.position = "bottom",
axis.text.x = element_blank(),
axis.title.x = element_blank())

## Warning: Removed 7 rows containing non-finite outside the scale range ('stat_boxplot()').

```

Life-history summary statistics



```

# save the plot using ggplot
ggsave(paste0("outputs/plots/manuscript_figs/memALL_summary_comparison_life_history_",
Sys.Date(), ".png"),
width=150, height=90, units="mm", dpi = 600)

```

Warning: Removed 7 rows containing non-finite outside the scale range ('stat_boxplot()').

Home range area (KDE) and monthly overlap

```

hr_summary_labs <- c("hr_area_50" = "KDE 50%",
"hr_area_75" = "KDE 75%",
"hr_area_95" = "KDE 95%",
"hr_monthly_overlap_ba" = "Monthly overlap")

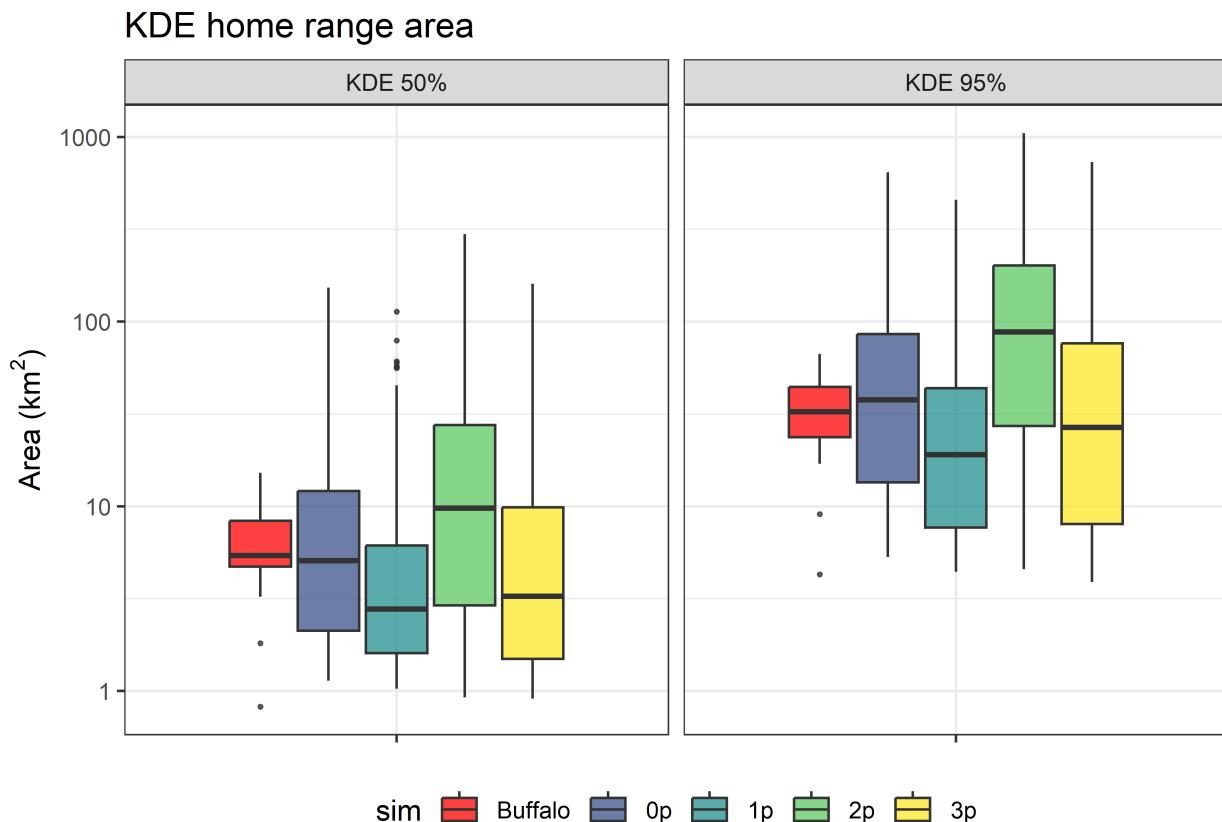
```

```

kde_hr_plot <- ggplot() +
  geom_boxplot(data = summaries_all_long %>%
    filter(category == "life_history" &
           str_detect(summary, "hr_area") &
           !str_detect(summary, "75")),
    aes(x = summary, y = value, fill = sim), alpha = 0.75,
    outlier.size = 0.5) +
  ggh4x::facet_grid2(scales = "free",
    cols = vars(summary),
    # independent = "y",
    remove_labels = "all",
    labeller = labeller(summary = hr_summary_labs)
  ) +
  ggtitle("KDE home range area") +
  scale_fill_manual(values = colors) +
  scale_y_log10(expression("Area (km" ^ 2 * "))) +
  theme_bw() +
  theme(legend.position = "bottom",
    axis.text.x = element_blank(),
    axis.title.x = element_blank())

```

kde_hr_plot

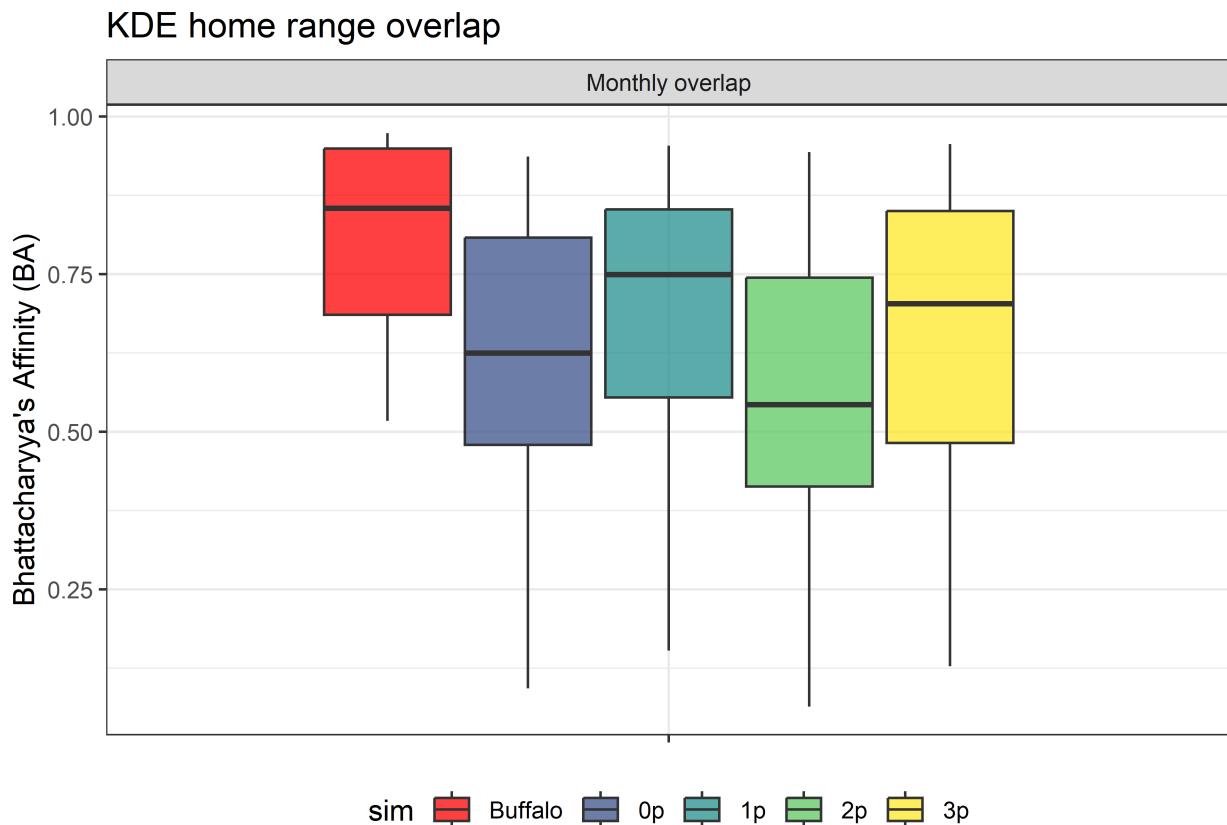


```

kde_hr_overlap_plot <- ggplot() +
  geom_boxplot(data = summaries_all_long %>%
    filter(category == "life_history" &
           str_detect(summary, "hr_monthly_overlap_ba")),
    aes(x = summary, y = value, fill = sim), alpha = 0.75,
    outlier.size = 0.5) +
  ggh4x::facet_grid2(scales = "free", cols = vars(summary), independent = "y",
    labeller = labeller(summary = hr_summary_labs))
  ) +
  ggtitle("KDE home range overlap") +
  scale_fill_manual(values = colors) +
  scale_y_continuous("Bhattacharyya's Affinity (BA)") +
  theme_bw() +
  theme(legend.position = "bottom",
    axis.text.x = element_blank(),
    axis.title.x = element_blank())

```

kde_hr_overlap_plot



Combining plots

```

ggarrange(kde_hr_plot +
  ggtitle("Home range area") +

```

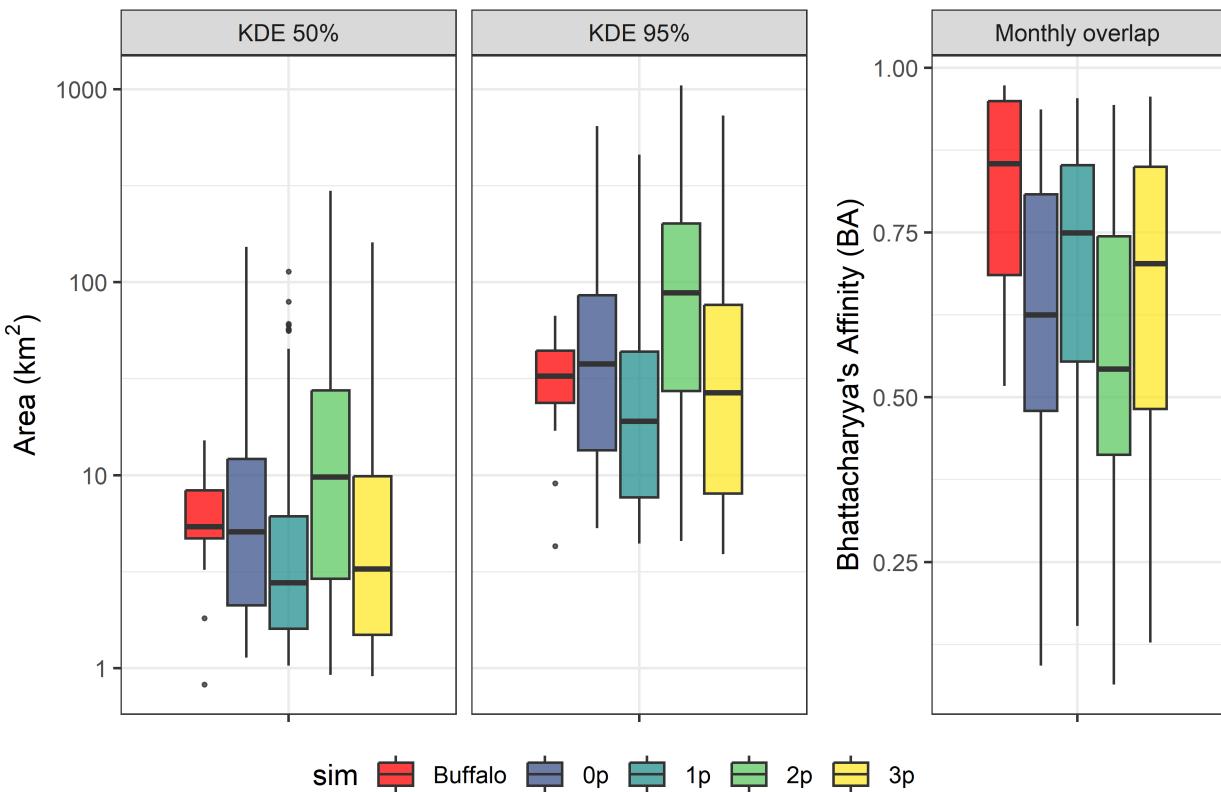
```

    theme(axis.title.x = element_blank()),

    kde_hr_overlap_plot +
      ggtitle("") +
      theme_bw() +
      theme(axis.title.x = element_blank(),
            axis.text.x = element_blank()),
  ncol = 2, nrow = 1,
  widths = c(2,1),
  legend = "bottom",
  common.legend = TRUE)

```

Home range area



```

ggsave(paste0("outputs/plots/manuscript_figs/memALL_kde_overlap_",
              Sys.Date(), ".png"),
       width=150, height=120, units="mm", dpi = 600)

```

Habitat summary statistics comparison

```

habitat_summaries_N0slope_plot <- ggplot() +
  geom_boxplot(data = summaries_all_long %>%
                filter(category == "habitat" &
                       str_detect(summary, "mean")) %>%
                filter(!str_detect(summary, "slope")),
                aes(x = summary, y = value, fill = sim), alpha = 0.75,

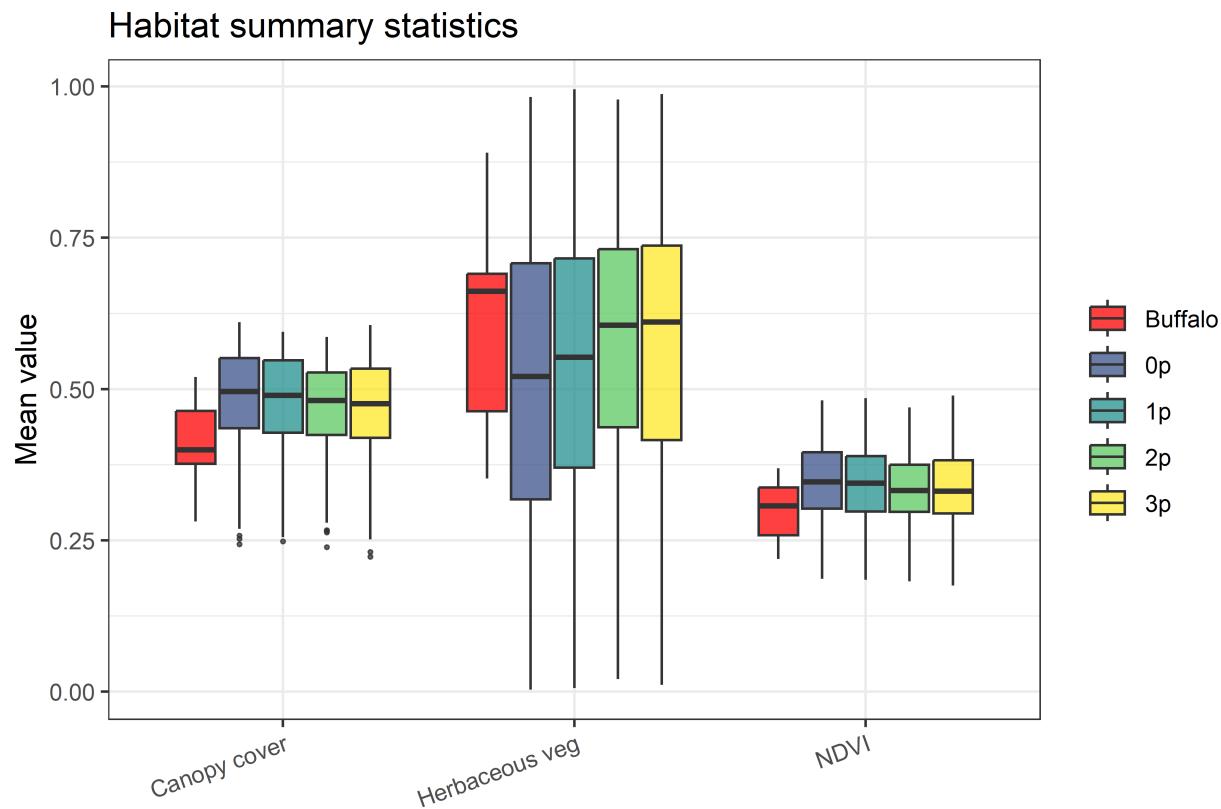
```

```

        outlier.size = 0.5) +
ggtitle("Habitat summary statistics") +
scale_fill_manual(values = colors, "") +
scale_x_discrete("", labels = c("Canopy cover", "Herbaceous veg", "NDVI")) +
scale_y_continuous("Mean value") +
theme_bw() +
theme(legend.position = "right",
axis.text.x = element_text(angle = 20, hjust = 1))

habitat_summaries_N0slope_plot

```



```

habitat_summaries_slope_plot <- ggplot() +
  geom_boxplot(data = summaries_all_long %>%
                filter(category == "habitat" &
                       str_detect(summary, "mean")) %>%
                filter(str_detect(summary, "slope")),
                aes(x = summary, y = value, fill = sim), alpha = 0.75,
                outlier.size = 0.5) +
  ggtitle(" ") +
  scale_fill_manual(values = colors, "") +
  scale_x_discrete("", labels = c("Slope")) +
  scale_y_continuous("Mean value") +
  theme_bw() +
  theme(legend.position = "right",

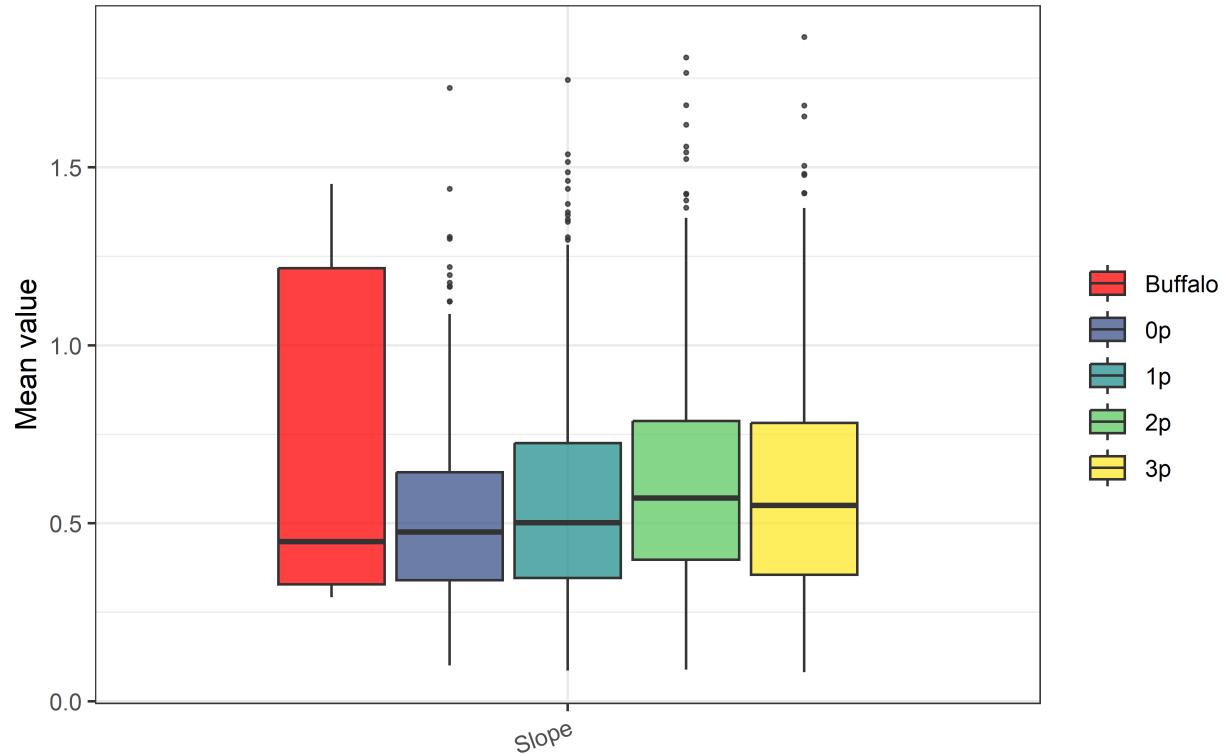
```

```

axis.text.x = element_text(angle = 20, hjust = 1))

habitat_summaries_slope_plot

```



Combining plots

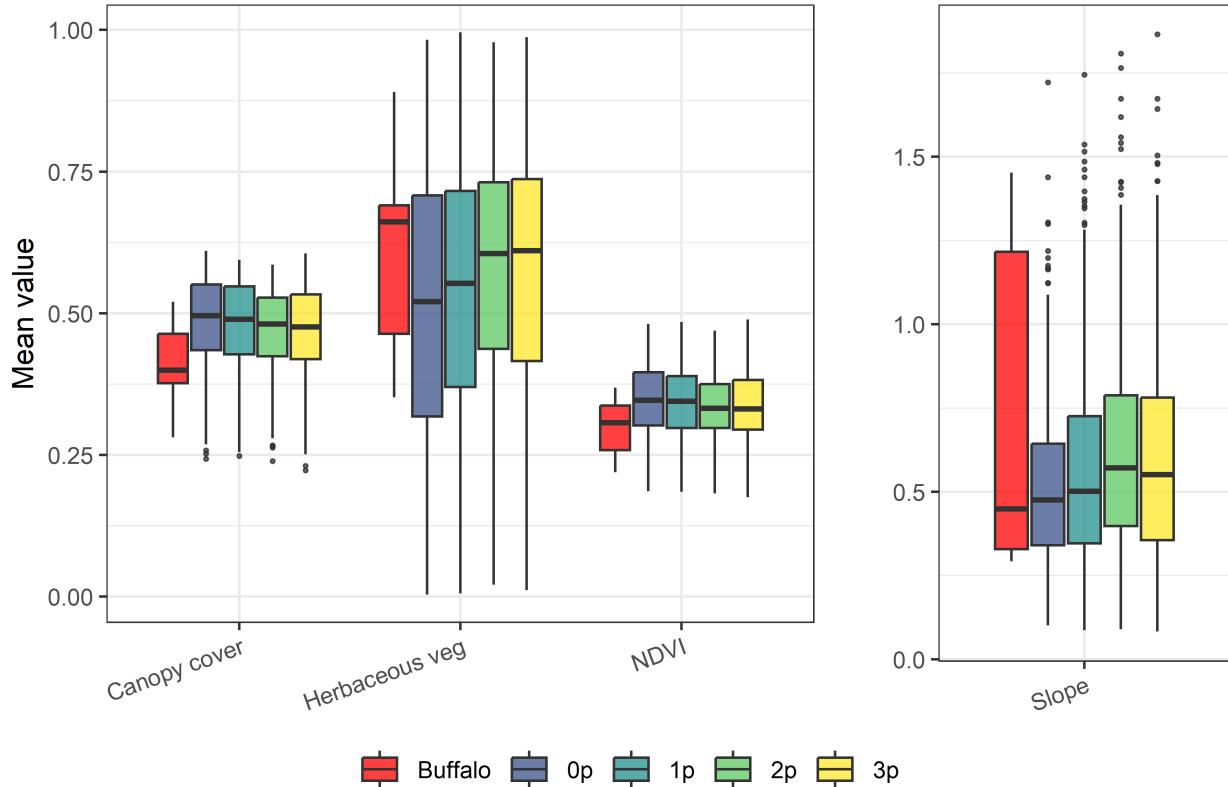
```

ggarrange(habitat_summaries_N0slope_plot + theme(axis.title.x = element_blank()),

          habitat_summaries_slope_plot + theme(axis.title.x = element_blank(),
                                                axis.title.y = element_blank()),
          ncol = 2, nrow = 1,
          widths = c(2,1),
          align = "v",
          legend = "bottom",
          common.legend = TRUE)

```

Habitat summary statistics



```
ggsave(paste0("outputs/plots/manuscript_figs/memALL_overall_habitat_overlap_",
              Sys.Date(), ".png"),
       width=150, height=120, units="mm", dpi = 600)
```

PCA of summary statistics

Similar to the approach of Abrahms et al. (2017) and Kays et al. (2023), we can use PCA to reduce the dimensionality of the summaries and assess the differences between the observed and simulated data in the latent space.

```
# remove some of the columns that should not be included in the PCA, such as id etc,
# and some that are covered by another summary, such as the monthly overlap with
# volume of intersection, as this information is largely contained in the Bhattacharyya's
# affinity overlap
# summaries_all_pca <- summaries_all %>%
#   dplyr::select(-traj, -id, -data, -sim, -residence_time_log,
#               -msd_log, -ba_overlap, -hr_monthly_overlap_vi) %>%
#   na.omit() %>%
#   prcomp(center = TRUE, scale. = TRUE)

summaries_all_pca <- summaries_all %>%
  dplyr::select(step_length_median,
                step_length_mean,
                gamma_shape,
                gamma_scale,
```

```

    vm_kappa,
    sinuosity,
    straightness,
    intensity_use,
    tac,
    residence_time,
    tot_dist,
    hr_area_50,
    hr_area_95,
    hr_monthly_overlap_ba,
    ndvi_mean,
    ndvi_sd,
    canopy_mean,
    canopy_sd,
    herby_mean,
    herby_sd,
    slope_mean,
    slope_sd
) %>%
na.omit() %>%
prcomp(center = TRUE, scale. = TRUE)

# summaries_all_pca

#PCA diagnostics
summary(summaries_all_pca)

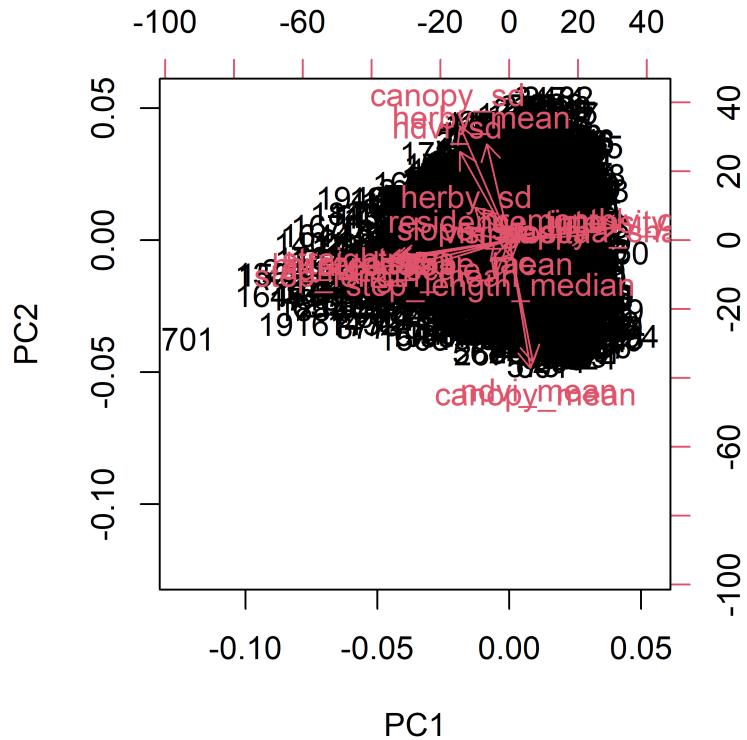
## Importance of components:
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10
## Standard deviation   2.4177 1.8589 1.4956 1.30693 1.17471 1.04737 1.01540 0.99208 0.92757 0.86182
## Proportion of Variance 0.2657 0.1571 0.1017 0.07764 0.06272 0.04986 0.04687 0.04474 0.03911 0.03376
## Cumulative Proportion 0.2657 0.4228 0.5244 0.60209 0.66482 0.71468 0.76155 0.80628 0.84539 0.87915
##                  PC17     PC18     PC19     PC20     PC21     PC22
## Standard deviation   0.37760 0.23851 0.18025 0.1556 0.09186 0.03776
## Proportion of Variance 0.00648 0.00259 0.00148 0.0011 0.00038 0.00006
## Cumulative Proportion 0.99439 0.99697 0.99845 0.9996 0.99994 1.00000
summaries_all_pca

## Standard deviations (1, ..., p=22):
## [1] 2.41774273 1.85892593 1.49563491 1.30693430 1.17470815 1.04737267 1.01539937 0.99207511 0.92757
## [14] 0.49857624 0.46897372 0.40445781 0.37759764 0.23850992 0.18024736 0.15561406 0.09186279 0.03775
## 
## Rotation (n x k) = (22 x 22):
##          PC1      PC2      PC3      PC4      PC5      PC6
## step_length_median -0.043610745 -0.14453303 0.416790542 -0.12096213 -0.26625254 0.149935648 0.0
## step_length_mean   -0.288534947 -0.11669106 0.336005862 -0.10712041 -0.14747165 0.303167530 -0.0
## gamma_shape        0.233403801 0.01944792 -0.238253850 0.22877698 -0.16007104 0.092013692 0.0
## gamma_scale        -0.302646278 -0.09631688 0.339096393 -0.15563730 -0.06457340 0.218754812 -0.0
## vm_kappa           0.018208272 0.01660181 0.182161840 -0.25558521 0.15910544 -0.535554831 -0.0
## sinuosity          0.040099225 0.01658025 -0.026794436 0.08099706 -0.02196747 -0.223326103 0.0
## straightness       -0.331677539 -0.06615495 -0.175614780 0.20960369 0.03874525 -0.222388174 0.0
## intensity_use      0.315304489 0.04749392 0.132835827 -0.13195630 -0.06399092 0.131231375 -0.0
## tac                0.007589292 -0.06731596 0.235708712 -0.18150705 -0.02943306 -0.429200589 0.0
## residence_time     -0.028518894 0.04939366 -0.264254452 0.09599451 0.08311065 0.378175518 -0.0

```

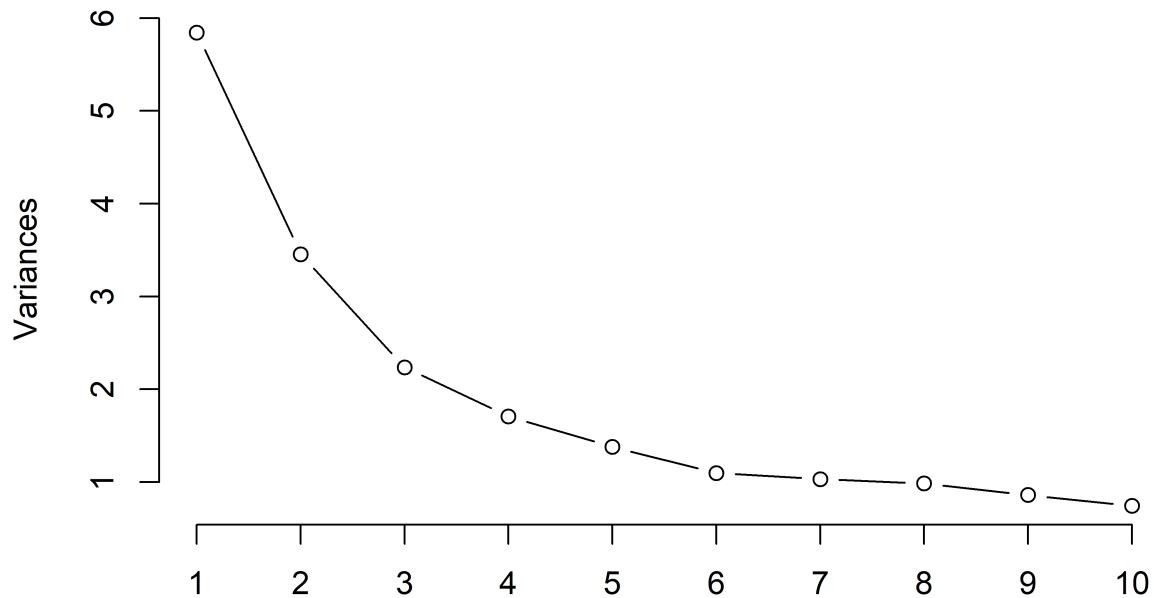
## tot_dist	-0.351615974	-0.07682222	-0.136714635	0.18545145	0.02511588	-0.159507479	0.0
## hr_area_50	-0.333370383	-0.07894714	-0.039791279	0.11660618	-0.03426936	0.006481809	-0.0
## hr_area_95	-0.365242858	-0.08282001	-0.052901649	0.12632525	-0.02490776	0.011507546	-0.0
## hr_monthly_overlap_ba	0.327093904	0.05397086	0.065294172	-0.04622356	-0.06657593	0.078261653	-0.0
## ndvi_mean	0.069851784	-0.46753582	-0.123038373	-0.15654387	0.25942183	0.031472490	0.0
## ndvi_sd	-0.145655443	0.33780383	-0.052505467	-0.29125715	0.36465795	0.082999119	0.0
## canopy_mean	0.062245597	-0.48886418	-0.113147766	-0.09246863	0.09608648	0.035135979	0.0
## canopy_sd	-0.145035694	0.43625662	0.003299475	-0.15735784	0.20456233	0.017283661	0.0
## herby_mean	-0.065954841	0.36768540	0.058244465	0.16882925	-0.44745228	-0.078875364	-0.0
## herby_sd	-0.100854879	0.12471861	-0.119518538	-0.31878549	0.25891329	0.214705499	0.0
## slope_mean	-0.055649944	-0.07974712	-0.370071902	-0.39643058	-0.44937142	-0.061489243	-0.0
## slope_sd	-0.112918445	0.02572922	-0.348114372	-0.47531459	-0.33317665	-0.037837836	0.0
##	PC11	PC12	PC13	PC14	PC15	PC16	
## step_length_median	0.326850279	-0.181271701	0.078440617	-0.009563826	0.063497515	-0.487581044	
## step_length_mean	0.093408210	0.049155245	0.052158925	-0.026053702	-0.133391227	0.431556424	
## gamma_shape	0.163071923	-0.299991898	-0.063390570	-0.049663820	-0.229812905	0.452241593	
## gamma_scale	0.007152455	0.124812870	0.067780654	-0.005555116	-0.030126261	0.227227305	
## vm_kappa	0.108214650	-0.162834758	-0.070401351	-0.024505892	-0.083987349	0.114691099	
## sinuosity	0.007189180	0.071543923	0.026092840	-0.002295316	0.002976754	0.007696557	
## straightness	0.132026506	0.067879299	0.460116165	0.042138876	0.179997562	-0.003586566	
## intensity_use	-0.271941117	-0.420034336	0.277853661	0.119682664	0.663955335	0.164823247	
## tac	-0.206921179	-0.118507169	0.019443574	0.009217104	-0.084749905	0.146294401	
## residence_time	0.155809813	-0.161407418	0.010115730	0.030059582	0.028220952	-0.078419057	
## tot_dist	0.100714675	0.015916943	0.459073331	0.055478903	0.142551047	0.066907469	
## hr_area_50	-0.382529998	-0.497641710	-0.171675709	-0.010937747	-0.130358713	-0.162351036	
## hr_area_95	-0.288549991	-0.335257570	-0.078560737	-0.014208562	-0.148889995	-0.029007537	
## hr_monthly_overlap_ba	-0.172832829	-0.094597124	0.650694441	-0.081898787	-0.584670087	-0.153349336	
## ndvi_mean	0.082933346	-0.082687451	0.006651462	-0.198700844	0.013008759	-0.088938083	
## ndvi_sd	0.197352219	-0.207362077	0.052013787	-0.233268345	-0.001584913	-0.052611950	
## canopy_mean	-0.107513190	0.079414476	-0.009094864	-0.415536760	0.083802877	-0.175376341	
## canopy_sd	0.195742212	-0.184871367	0.032781814	-0.242029520	-0.005521247	-0.043903308	
## herby_mean	-0.218446842	0.188547525	-0.043600824	-0.442907031	0.128948622	-0.211661844	
## herby_sd	-0.515290493	0.331149355	0.092648136	0.037640172	-0.005771972	0.056121948	
## slope_mean	0.075365176	-0.038910898	0.048769203	-0.402847269	0.060351312	0.240021434	
## slope_sd	0.057740485	-0.009652875	-0.030343441	0.532838629	-0.085590560	-0.218654149	
##	PC20	PC21	PC22				
## step_length_median	-0.002185974	-0.001285664	0.022129748938				
## step_length_mean	0.020995281	-0.027158498	-0.617237694828				
## gamma_shape	0.013079995	-0.029849929	0.232276263824				
## gamma_scale	0.018480059	-0.056432498	0.750264765647				
## vm_kappa	-0.003114502	0.002379166	-0.001413285580				
## sinuosity	0.003678489	0.003492837	-0.000008426762				
## straightness	0.003557722	-0.674634190	-0.023923582437				
## intensity_use	-0.009423794	-0.027049419	-0.015361853299				
## tac	-0.002247738	-0.002261910	0.003979397524				
## residence_time	-0.002950762	-0.003143092	0.000780913529				
## tot_dist	0.021600396	0.724468327	0.023403643970				
## hr_area_50	0.092903490	0.037062222	-0.014366170173				
## hr_area_95	-0.119623702	-0.105393992	0.006202362122				
## hr_monthly_overlap_ba	-0.007145401	-0.016511544	-0.004810403601				
## ndvi_mean	0.727496881	-0.016329275	-0.000833356725				
## ndvi_sd	-0.015232674	-0.008893513	-0.003478077428				
## canopy_mean	-0.585632399	0.030100457	0.001971091413				
## canopy_sd	-0.120002263	0.011483006	0.003830996344				

```
## herby_mean          0.285862219  0.013384192 -0.000664720940  
## herby_sd            0.080153244 -0.009531193 -0.000286864188  
## slope_mean          -0.006873446 -0.004876558 -0.000823142733  
## slope_sd           -0.020155690  0.001847557  0.001836341939  
  
biplot(summaries_all_pca)
```



```
screeplot(summaries_all_pca, type = "1")
```

summaries_all_pca



```

# proportion of variance explained
prop_var <- summaries_all_pca$sdev^2 / sum(summaries_all_pca$sdev^2)

# to create of vector of the simulation labels
# this should remove the same columns as above, besides sim
sims <- summaries_all %>%
  dplyr::select(-traj, -id, -data, -residence_time_log, -msd_log,
    -ba_overlap, -hr_monthly_overlap_vi) %>%
  na.omit() %>%
  pull(sim)

# Create a data frame for ggplot
pca_data <- data.frame(sim = sims, summaries_all_pca$x)

# reorder levels for the legend
pca_data$sim <- factor(pca_data$sim, levels = c("obs", "0p", "1p", "2p", "3p"),
  labels = c("Buffalo", "0p", "1p", "2p", "3p"))

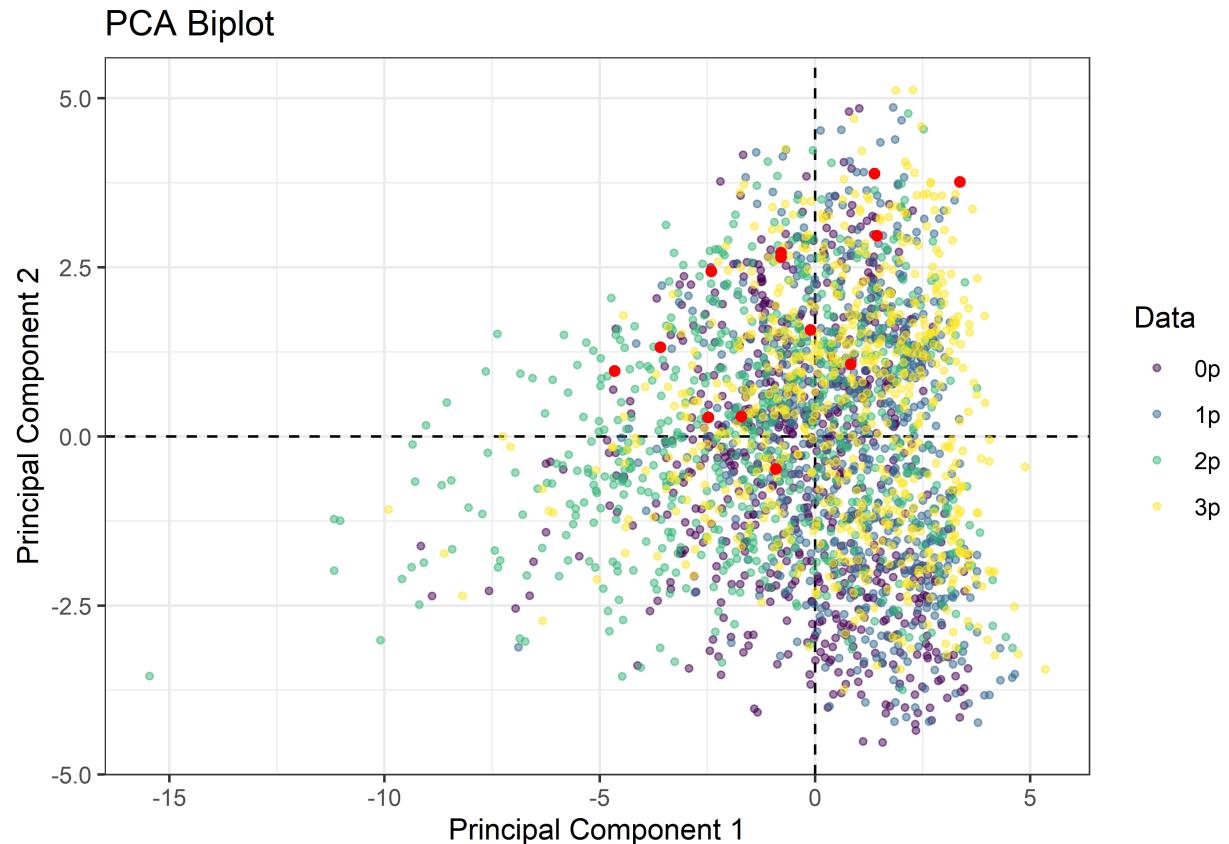
# Biplot
ggplot() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  geom_point(data = pca_data %>% filter(sims != "obs"),
    aes(PC1, PC2, colour = sim),
    size = 1, alpha = 0.5) +
  geom_point(data = pca_data %>% filter(sims == "obs"),

```

```

aes(PC1, PC2),
colour = "red") +
scale_colour_viridis_d("Data") +
ggtitle("PCA Biplot") +
xlab("Principal Component 1") +
ylab("Principal Component 2") +
theme_bw()

```



We see that the simulated and observed summaries overlap in the PCA space, and there are few differences between the simulated trajectories from different models, suggesting that the simulations are largely replicating the trends observed in the data, given our set of summary statistics.

We can add a few more plotting elements for further comparison.

```

# Function to compute convex hull and return a data frame
get_hull_coordinates <- function(data, group) {
  points <- data[data$sim == group,]
  hull <- chull(points$PC1, points$PC2)
  return(data.frame(points[hull, ]))
}

# Compute hulls for each group
hulls <- do.call(rbind, lapply(unique(pca_data$sim), get_hull_coordinates, data = pca_data))

```

Add density to the x and y axes.

```

# Generate Viridis colors for three groups
viridis_colors <- viridis::viridis(4)
# Create a named vector of colors for all groups
colors <- setNames(c('red', viridis_colors), c("Buffalo", "0p", "1p", "2p", "3p"))

# Biplot with polygons
pca_plot <- ggplot() +
  geom_point(data = pca_data,
             aes(PC1, PC2, color = sim),
             alpha = 0.5) +
  geom_polygon(data = hulls %>% filter(sim == "Buffalo"),
               aes(x = PC1, y = PC2, group = sim),
               alpha = 0.3, fill = "red") +
  theme_bw() +
  ggtitle("PCA of all summary statistics") +
  xlab(paste0("PC1: ", round(prop_var[1]*100, 1), "% expl. variance")) +
  ylab(paste0("PC2: ", round(prop_var[2]*100, 1), "% expl. variance")) +
  scale_color_manual("Data", values = colors) +
  # scale_y_continuous(limits = c(-5.5, 5.5)) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_vline(xintercept = 0, linetype = "dashed") +
  theme(legend.position = "bottom")

# pca_plot

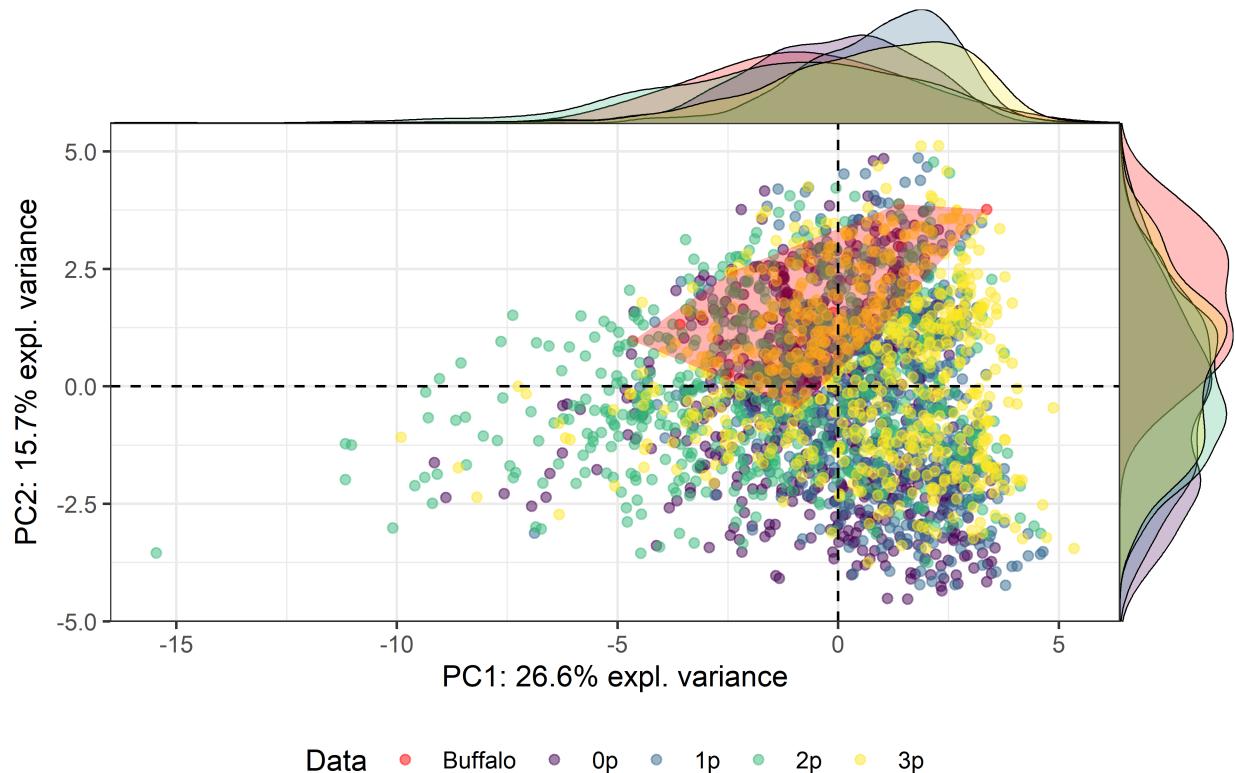
# Add marginal density plots
xdens <-
  axis_canvas(pca_plot, axis = "x") +
  geom_density(data = pca_data, aes(x = PC1, fill = sim), size = 0.25, alpha = 0.25) +
  scale_fill_manual(values = colors)

ydens <-
  axis_canvas(pca_plot, axis = "y", coord_flip = TRUE) +
  geom_density(data = pca_data, aes(x = PC2, fill = sim), size = 0.25, alpha = 0.25) +
  scale_fill_manual(values = colors) +
  coord_flip()

pca_plot %>%
  insert_xaxis_grob(xdens, grid::unit(1.5, "cm"), position = "top") %>%
  insert_yaxis_grob(ydens, grid::unit(1.5, "cm"), position = "right") %>%
  ggdraw()

```

PCA of all summary statistics



```
# save the plot using ggplot
ggsave(paste0("outputs/plots/manuscript_figs/memALL_PCA_all_summaries_",
              Sys.Date(), ".png"),
       width=150, height=120, units="mm", dpi = 600)
```

References

- Abrahms, Briana, Dana P Seidel, Eric Dougherty, Elliott L Hazen, Steven J Bograd, Alan M Wilson, J Weldon McNutt, et al. 2017. “Suite of Simple Metrics Reveals Common Movement Syndromes Across Vertebrate Taxa.” *Movement Ecology* 5 (1): 12. <https://doi.org/10.1186/s40462-017-0104-2>.
- Kays, Roland, Ben Hirsch, Damien Caillaud, Rafael Mares, Shauhin Alavi, Rasmus Worsøe Havmøller, and Margaret Crofoot. 2023. “Multi-Scale Movement Syndromes for Comparative Analyses of Animal Movement Patterns.” *Movement Ecology* 11 (1): 61. <https://doi.org/10.1186/s40462-022-00365-y>.

Session info

```
sessionInfo()

## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_New Zealand.utf8  LC_CTYPE=English_New Zealand.utf8    LC_MONETARY=English_New Zealand.UTF-8
## [4] LC_NUMERIC=C                           LC_TIME=English_New Zealand.utf8
```

```

## [4] LC_NUMERIC=C                               LC_TIME=English_New Zealand.utf8
##
## attached base packages:
## [1] stats      graphics   grDevices  utils      datasets   methods    base
##
## other attached packages:
## [1] scales_1.3.0     viridis_0.6.2    viridisLite_0.4.2 ggpubr_0.6.0    cowplot_1.1.1  ggExti...
## [9] ggh4x_0.2.6      lubridate_1.9.2 forcats_1.0.0     stringr_1.5.0    dplyr_1.1.2    purrr_...
## [17] tibble_3.2.1     ggplot2_3.5.0  tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] sass_0.4.5       jsonlite_1.8.4   bit64_4.0.5      vroom_1.6.1     splines_4.2.1 carDa...
## [9] highr_0.10      yaml_2.3.7      pillar_1.9.0     backports_1.4.1 lattice_0.21-8 glue_...
## [17] ggsignif_0.6.4  colorspace_2.1-0 htmltools_0.5.5  httpuv_1.6.9   Matrix_1.6-5  pkgco...
## [25] later_1.3.0     tzdb_0.3.0      timechange_0.2.0 mgcv_1.8-42    farver_2.1.1  genera...
## [33] cachem_1.0.7    withr_2.5.0     cli_3.6.1       magrittr_2.0.3  crayon_1.5.2 mime_...
## [41] nlme_3.1-162    rstatix_0.7.2   textshaping_0.3.6 tools_4.2.1    hms_1.1.3    lifec...
## [49] compiler_4.2.1   systemfonts_1.0.4 rlang_1.1.0     grid_4.2.1     rstudioapi_0.14 miniU...
## [57] gtable_0.3.3    codetools_0.2-19 abind_1.4-5     R6_2.5.1      gridExtra_2.3 knitr_...
## [65] utf8_1.2.3      ragg_1.2.5      stringi_1.7.12  parallel_4.2.1 Rcpp_1.0.10  vctrs_...

```