Data exploration - Principal component analysis

Anonymous

2021-10-19

Contents

Summary

The following script explores the data used for the analyses presented in "title". We plot the distributions of the raw and transformed variables for predators and prey. We then use a principal component analysis to explore the correlations between the hunting behaviors. We further investigate the distributions of player experience, and look at the correlations between player experience metrics.

1. Prepare the data

1.1 Load the required packages

```
using Pkg, Weave, CSV, DataFrames, Plots
```

1.2 Import the data

```julia
data = DataFrame(CSV.File(joinpath(path, "03_final-data",
                                         "03_final-data_2021",
                                         "final-data.csv")))
```

2. Inspect the data

2.1 Distribution of predator behavior

In this section, we will inspect the distribution of the different predator behaviors. We provide the distribution of the raw variables and their transformations.

The transformations include :

```
- square root
- log
```

2.1.1 Raw variables

```
# Predator behavior
a = histogram(data.pred_speed,
              bins = 40,
              label = "speed")
b = histogram(data.pred_amount_tiles_visited,
              bins = 40,
              label = "space covered")
c = histogram(data.ambush_time_close,
              bins = 40,
              label = "ambush time")
d = histogram(data.latency_1st_capture,
              bins = 40,
              label = "latency 1st capture")

e = histogram(data.total_chase_duration,
              bins = 40,
              label = "chase duration")

plot(a, b, c, d, layout = 4)
```
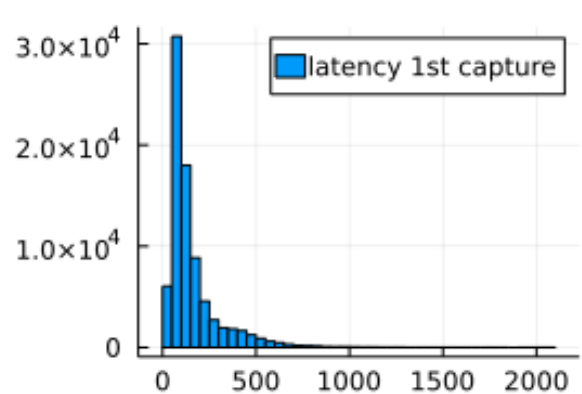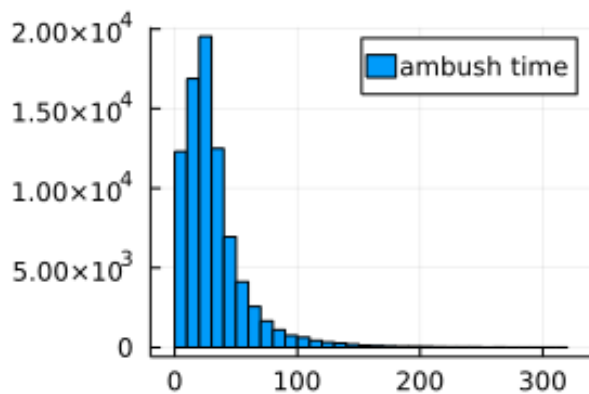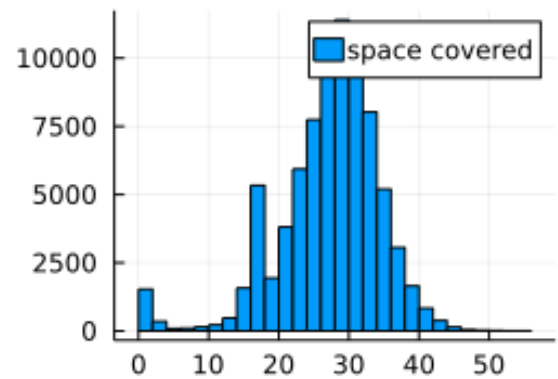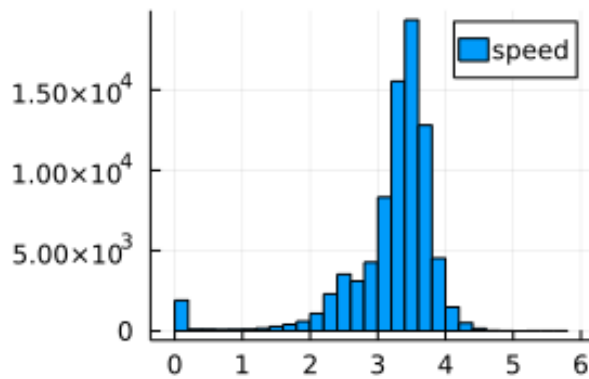
2.1.2 Square root-transformed variables

```
# Square root of predator behavior
a1 = histogram(sqrt.(data.pred_speed),
               bins = 40,
               label = "(speed)²")
b1 = histogram(sqrt.(data.pred_amount_tiles_visited),
               bins = 40,
               label = "(space covered)²")
c1 = histogram(sqrt.(data.ambush_time_close),
               bins = 40,
               label = "(ambush time)²")
d1 = histogram(sqrt.(data.latency_1st_capture),
               bins = 40,
               label = "(latency 1st capture)².")

plot(a1, b1, c1, d1, layout = 4)
```
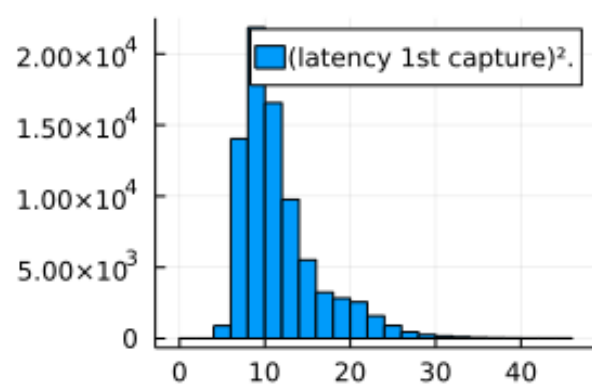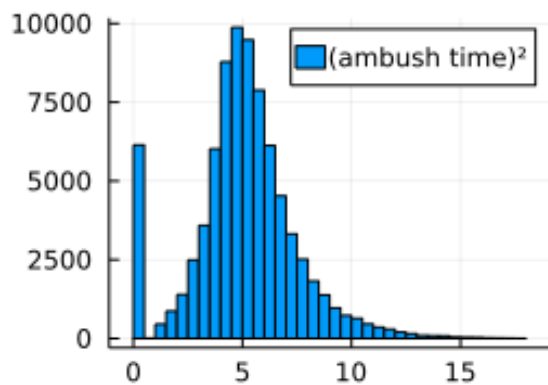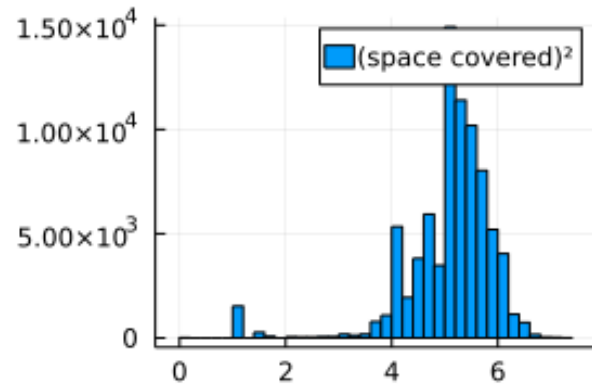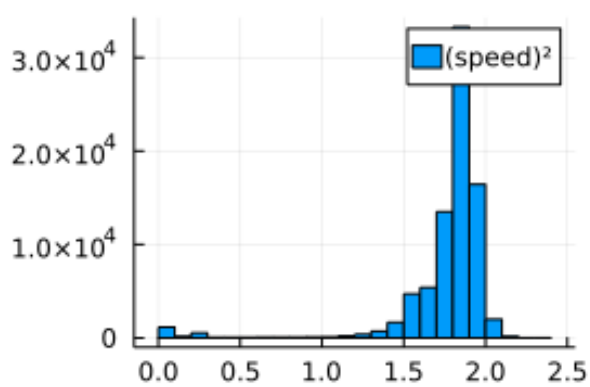
2.1.3 Log-transformed variables

```
# Log of predator behavior
a2 = histogram(log.(data.pred_speed),
               bins = 40,
               label = "log(speed)")
b2 = histogram(log.(data.pred_amount_tiles_visited),
               bins = 40,
               label = "log(space covered)")
c2 = histogram(log.(data.ambush_time_close),
               bins = 40,
               label = "log(ambush time)")
d2 = histogram(log.(data.latency_1st_capture), bins = 40,
               label = "log(latency 1st capture)")

plot(a2, b2, c2, d2, layout = 4)
```
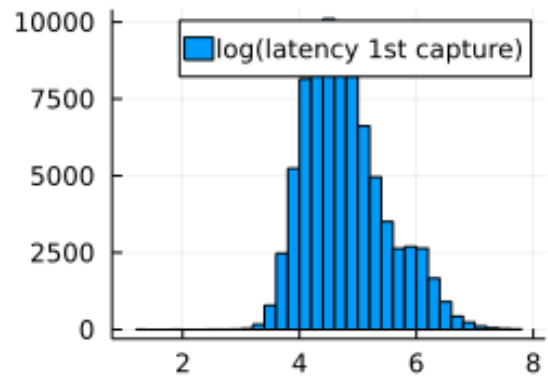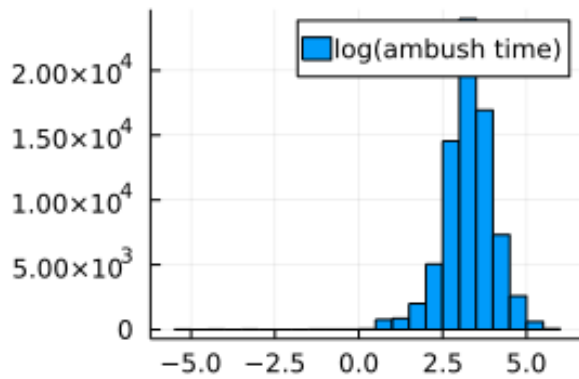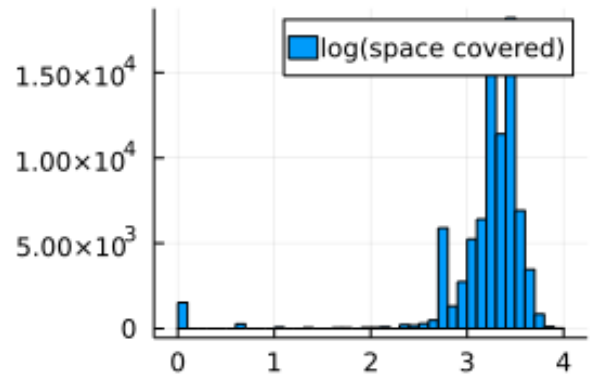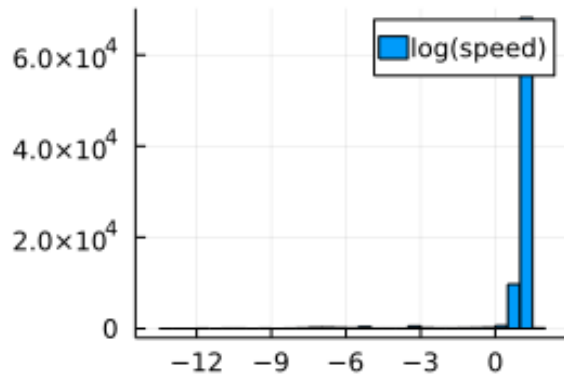
2.2 Distribution of prey behavior

We now inspect the distribution of the prey behaviors. The prey behaviors were calculated using the average values of the 4 prey within a match. We provide the distribution of the raw variables and their transformations.

The transformations include :

- square root
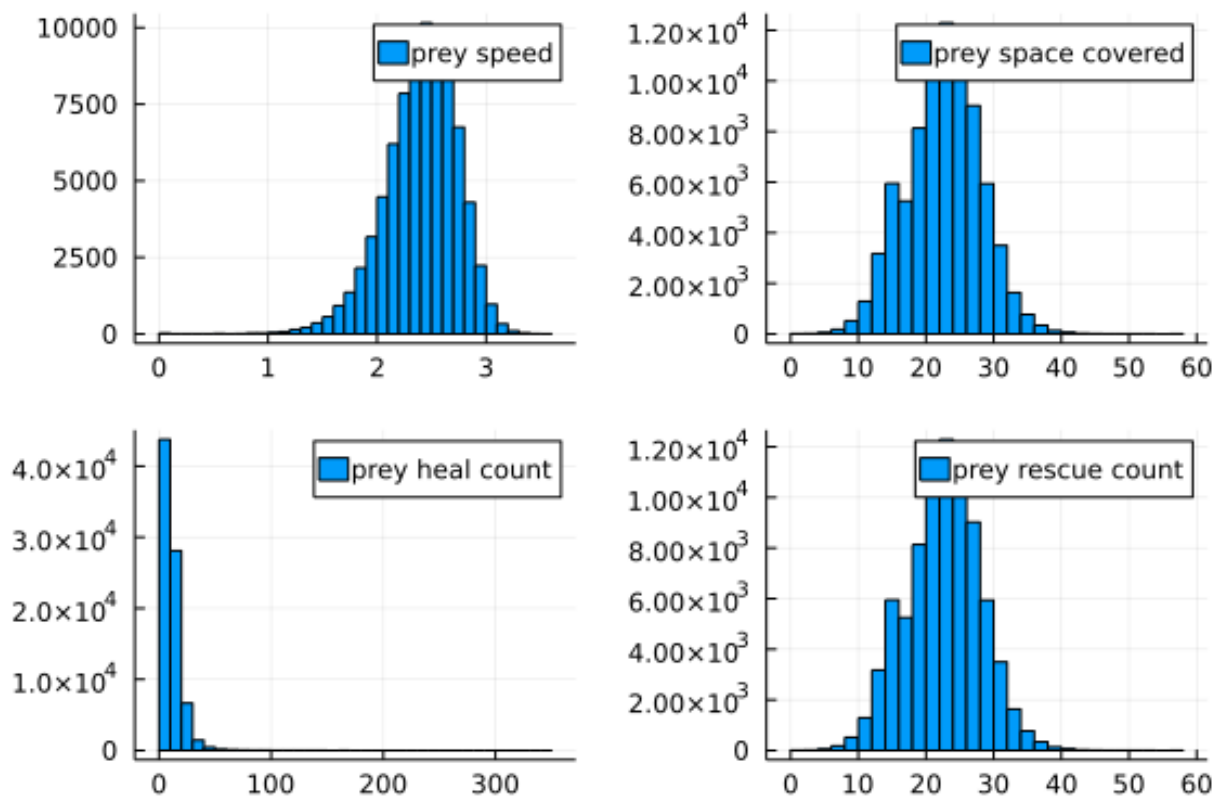- log

2.2.1 Raw variables

```
# Prey behavior
f = histogram(data.prey_avg_speed,
              bins = 40,
              label = "prey speed")
g = histogram(data.prey_avg_amount_tiles_visited,
              bins = 40,
              label = "prey space covered")

h = histogram(data.prey_total_heal_count,
              bins = 40,
              label = "prey heal count")

i = histogram(data.prey_avg_amount_tiles_visited,
              bins = 40,
              label = "prey rescue count")

plot(f, g, h, i, layout = 4)
```

2.2.2 Square root-transformed variables

```julia
# Prey behavior
f1 = histogram(sqrt.(data.prey_avg_speed),
               bins = 40,
               label = "(prey speed)²")
g1 = histogram(sqrt.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "(prey space covered)²")

h1 = histogram(sqrt.(data.prey_total_heal_count),
               bins = 40,
               label = "(prey heal count)²")

i1 = histogram(sqrt.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "(prey rescue count)²")

plot(f1, g1, h1, i1, layout = 4)
```
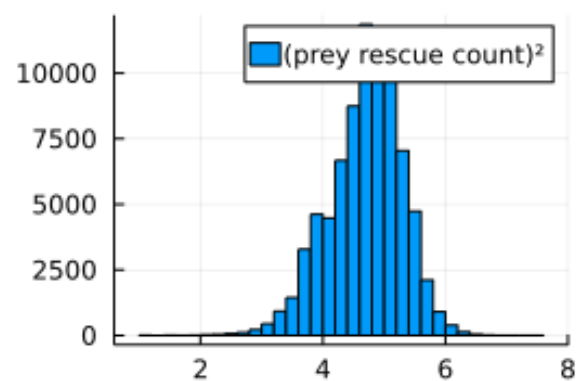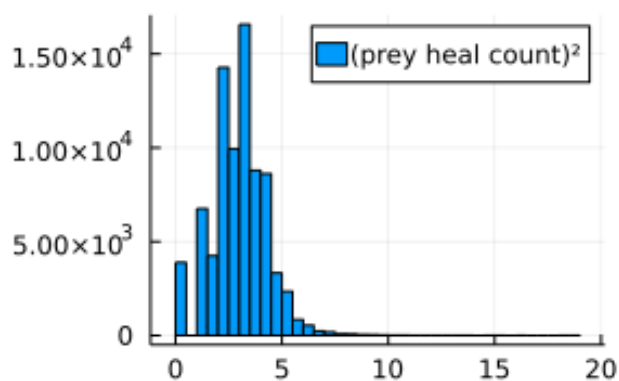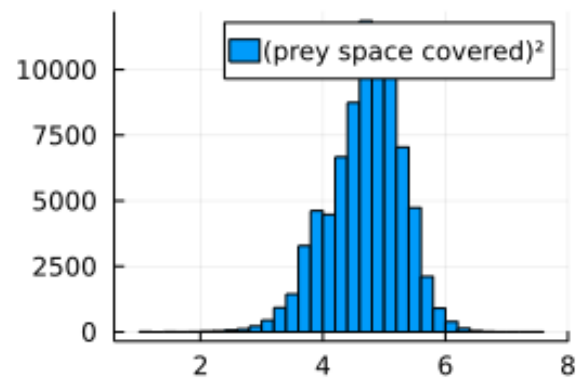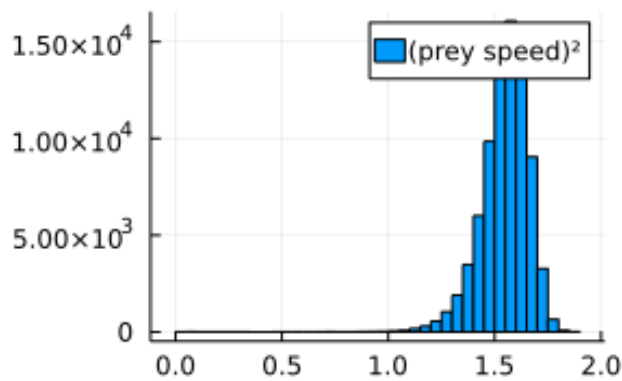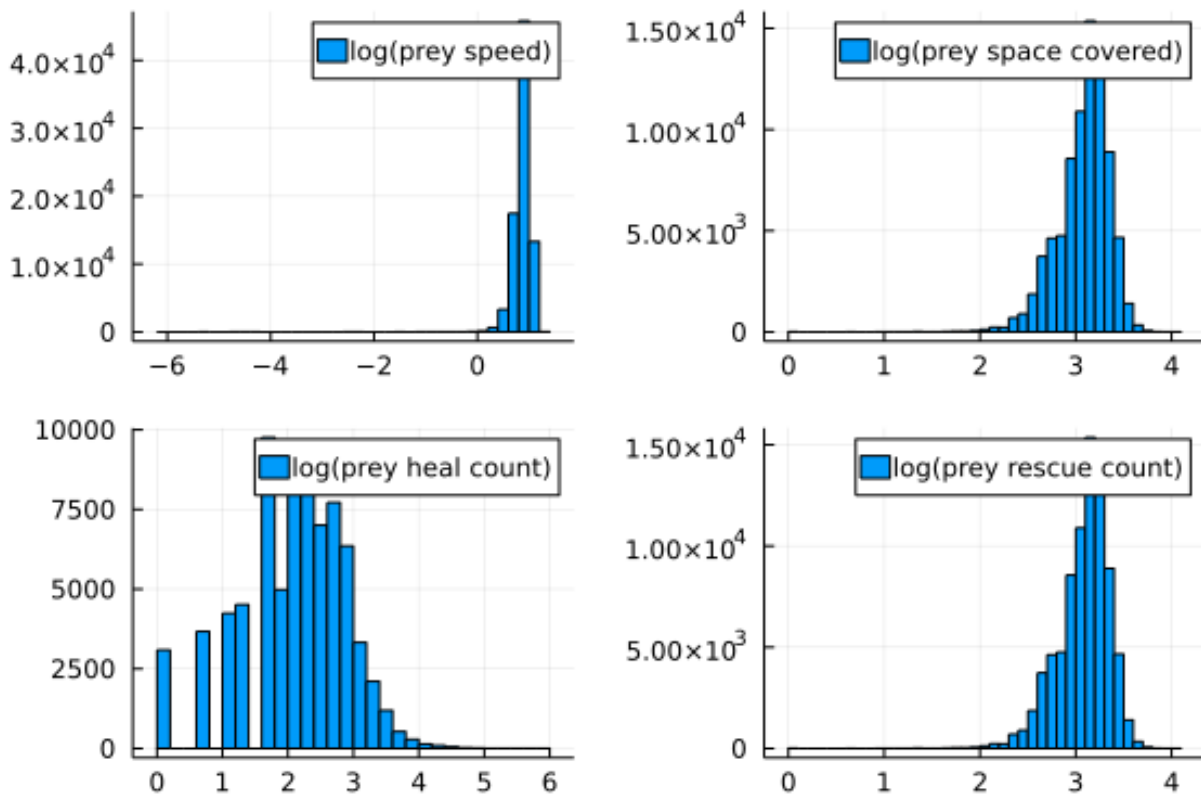
2.2.3 Log-transformed variables

```
# Prey behavior
f2 = histogram(log.(data.prey_avg_speed),
               bins = 40,
               label = "log(prey speed)")
g2 = histogram(log.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "log(prey space covered)")

h2 = histogram(log.(data.prey_total_heal_count),
               bins = 40,
               label = "log(prey heal count)")

i2 = histogram(log.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "log(prey rescue count)")

plot(f2, g2, h2, i2, layout = 4)
```


3. Transformed-variables selected for the analyse