

Data exploration - Principal component analysis

Anonymous

2021-10-19

Contents

Summary	2
1. Prepare the data	3
1.1 Variable definitions	3
1.2 Load the required packages	3
1.3 Import the data	3
2. Inspect the data	3
2.1 Distribution of predator behavior	3
2.1.1 Raw variables	3
2.1.2 Square root-transformed variables	5
2.1.3 Log-transformed variables	6
2.2 Distribution of prey behavior	7
2.2.1 Raw variables	7
2.2.2 Square root-transformed variables	8
2.2.3 Log-transformed variables	9
2.3 Distribution of player experience	10
3. Transformed-variables selected for the analyses	11

Summary

The following script explores the data used for the analyses presented in “title”. We plot the distributions of the raw and transformed (log and square root) variables for predators and prey. We then use a principal component analysis to explore the covariances between the hunting behaviors. We further investigate the distributions of player experience, and look at the correlations between player experience variables. The data is structured like a longitudinal dataset with repeated individual measurements. The prey behaviors were calculated using the average or the sum of values of the 4 prey within a match. There is thus one observations per match for each predator player.

1. Prepare the data

1.1 Variable definitions

The predator behaviors :

- average speed (m/s)
- space covered (amount of tiles visited)
- time spent ambushing (s)
- latency before the first prey capture (s)

The prey behaviors :

- average speed (m/s)
- average space covered (amount of tiles visited)
- total amount of times the prey healed another player
- total amount of times prey rescued another player

The player experience metrics :

- cumulative sum of all matches (as a predator and as a prey)
- total matches played (as a predator and as a prey)
- cumulative sum of matches as the predator
- total matches played as the predator

1.2 Load the required packages

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

1.3 Import the data

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

2. Inspect the data

In this section, we will first plot the distributions of the different predator behaviors (in red). We follow by plotting the distributions of the prey behaviors.

2.1 Distribution of predator behavior

2.1.1 Raw variables

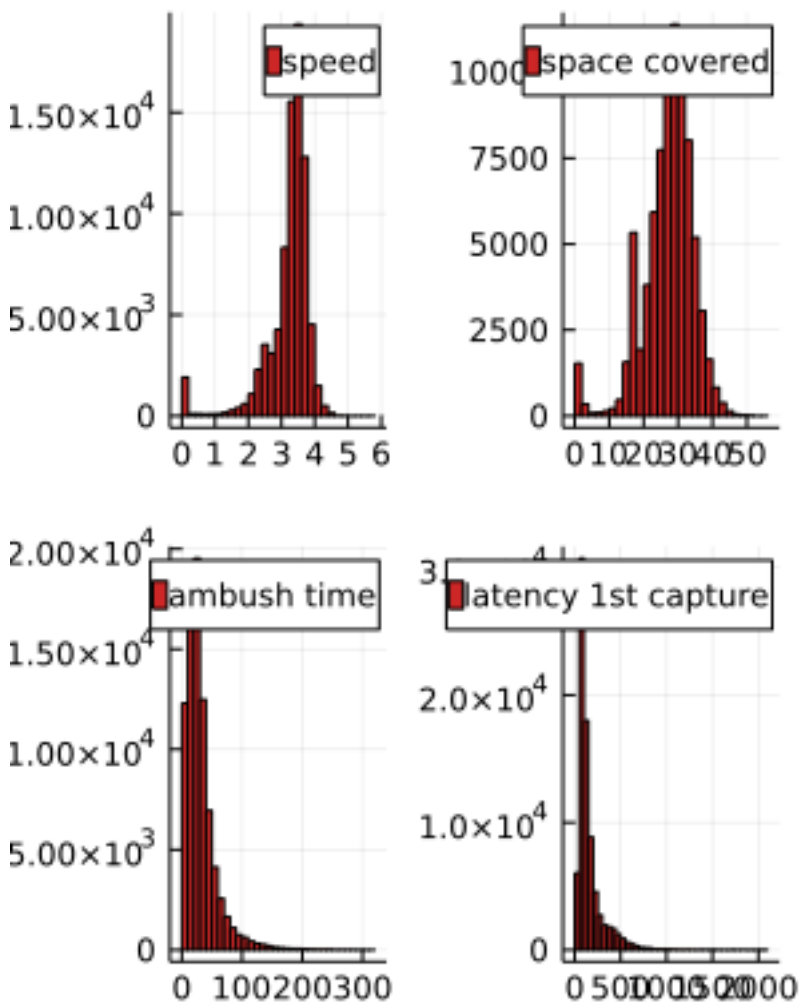
```
a = histogram(data.pred_speed,  
              bins = 40,  
              label = "speed",  
              color = :firebrick3)  
b = histogram(data.pred_amount_tiles_visited,  
              bins = 40,
```

```

        label = "space covered",
        color = :firebrick3)
c = histogram(data.ambush_time_close,
        bins = 40,
        label = "ambush time",
        color = :firebrick3)
d = histogram(data.latency_1st_capture,
        bins = 40,
        label = "latency 1st capture",
        color = :firebrick3)
e = histogram(data.total_chase_duration,
        bins = 40,
        label = "chase duration",
        color = :firebrick3)

plot(a, b, c, d, layout = 4, size = (300, 400))

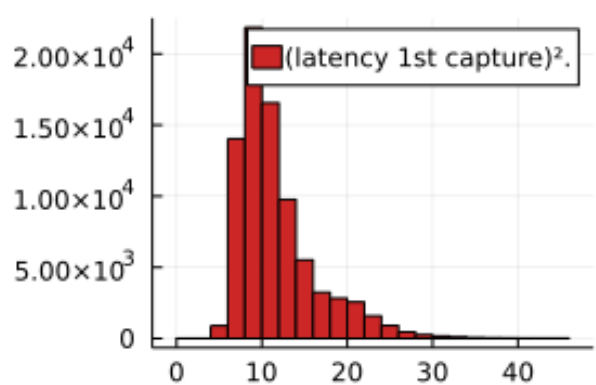
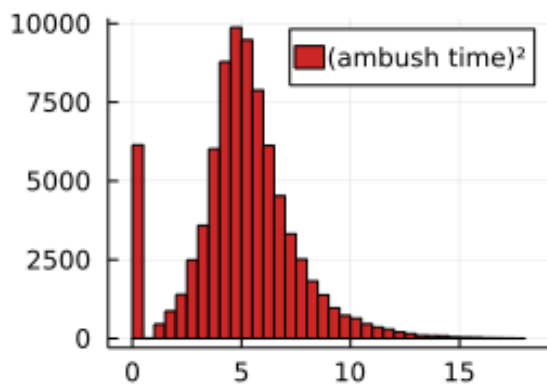
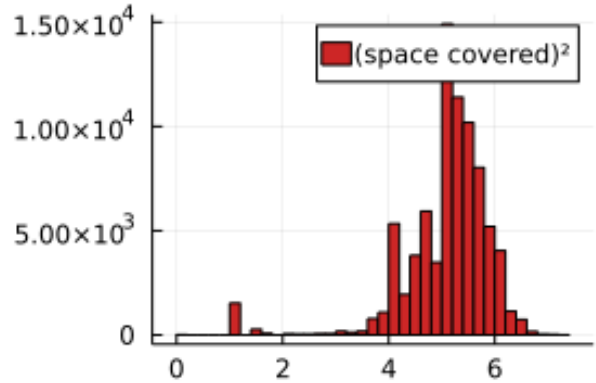
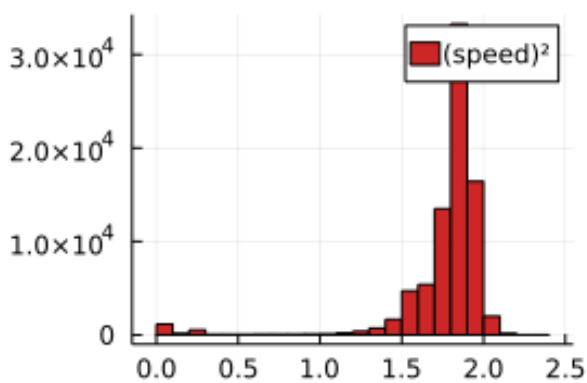
```



2.1.2 Square root-transformed variables

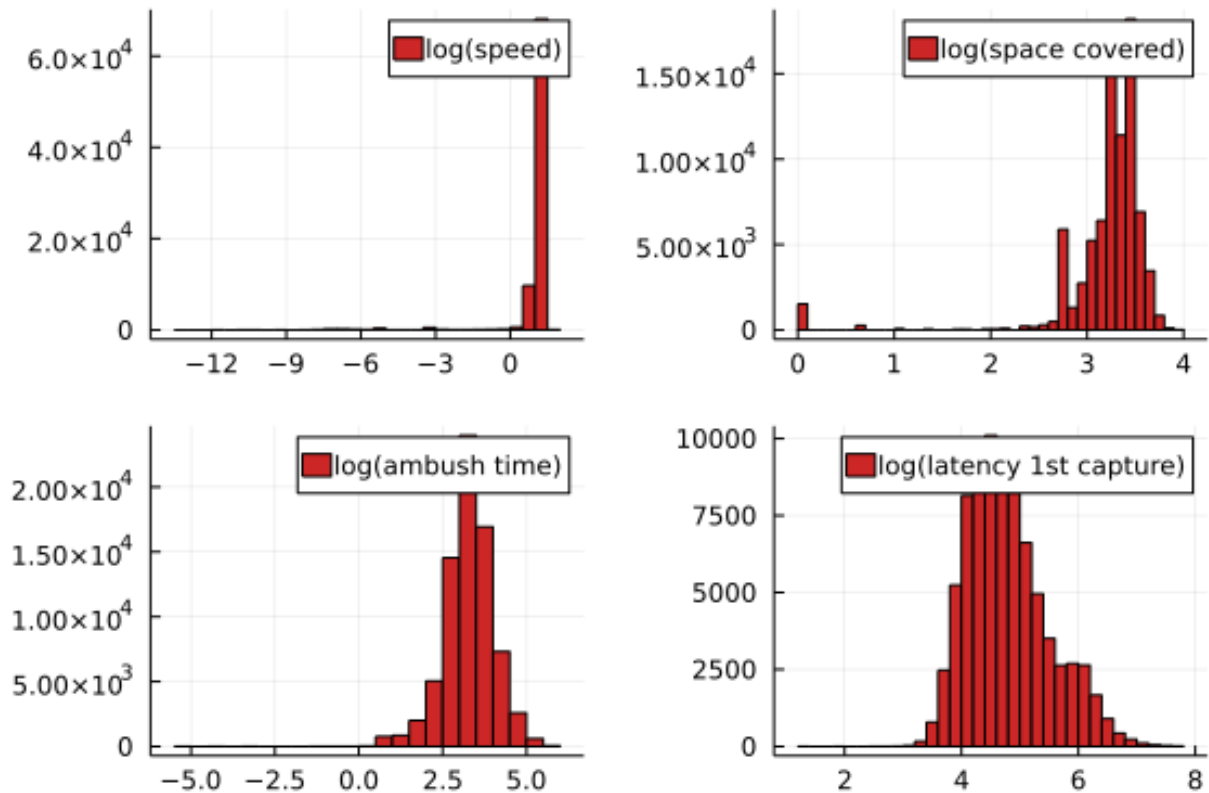
```
a1 = histogram(sqrt.(data.pred_speed),
               bins = 40,
               label = "(speed)2",
               color = :firebrick3)
b1 = histogram(sqrt.(data.pred_amount_tiles_visited),
               bins = 40,
               label = "(space covered)2",
               color = :firebrick3)
c1 = histogram(sqrt.(data.ambush_time_close),
               bins = 40,
               label = "(ambush time)2",
               color = :firebrick3)
d1 = histogram(sqrt.(data.latency_1st_capture),
               bins = 40,
               label = "(latency 1st capture)2.",
               color = :firebrick3)

plot(a1, b1, c1, d1, layout = 4)
```



2.1.3 Log-transformed variables

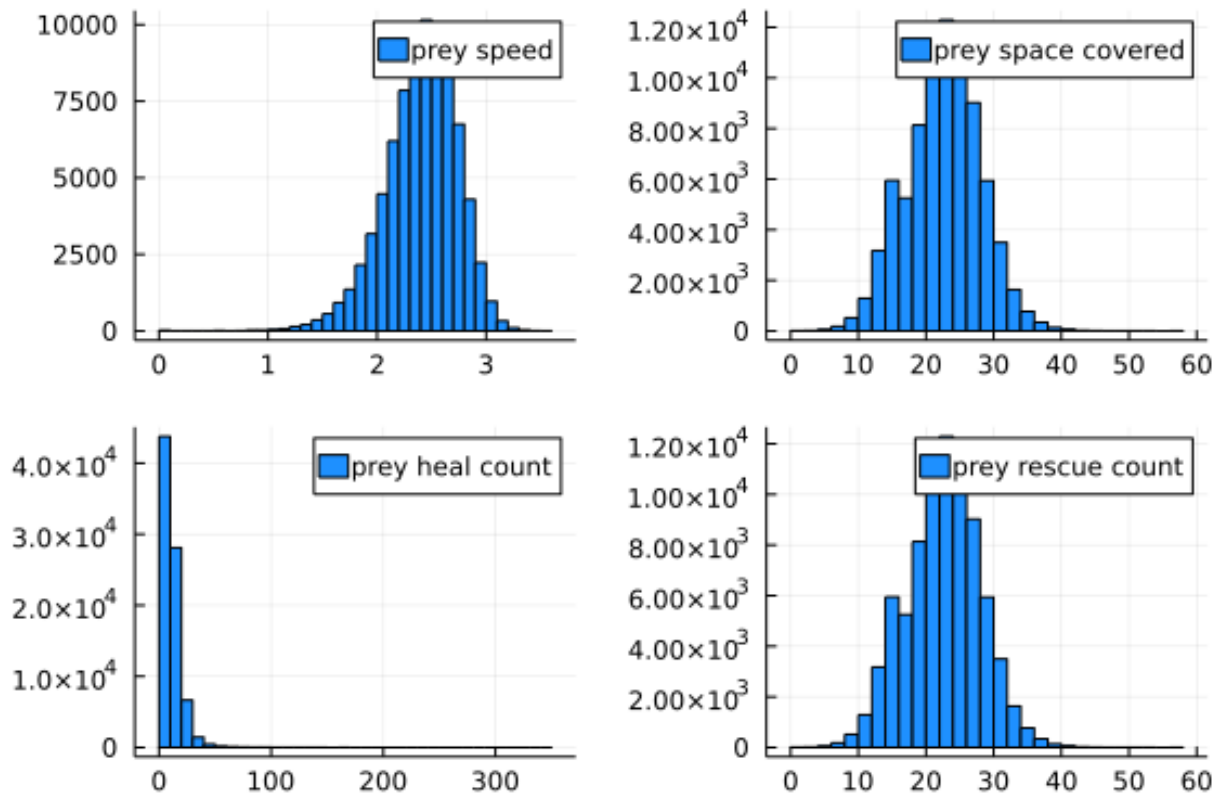
```
a2 = histogram(log.(data.pred_speed),  
               bins = 40,  
               label = "log(speed)",  
               color = :firebrick3)  
b2 = histogram(log.(data.pred_amount_tiles_visited),  
               bins = 40,  
               label = "log(space covered)",  
               color = :firebrick3)  
c2 = histogram(log.(data.ambush_time_close),  
               bins = 40,  
               label = "log(ambush time)",  
               color = :firebrick3)  
d2 = histogram(log.(data.latency_1st_capture), bins = 40,  
               label = "log(latency 1st capture)",  
               color = :firebrick3)  
  
plot(a2, b2, c2, d2, layout = 4)
```



2.2 Distribution of prey behavior

2.2.1 Raw variables

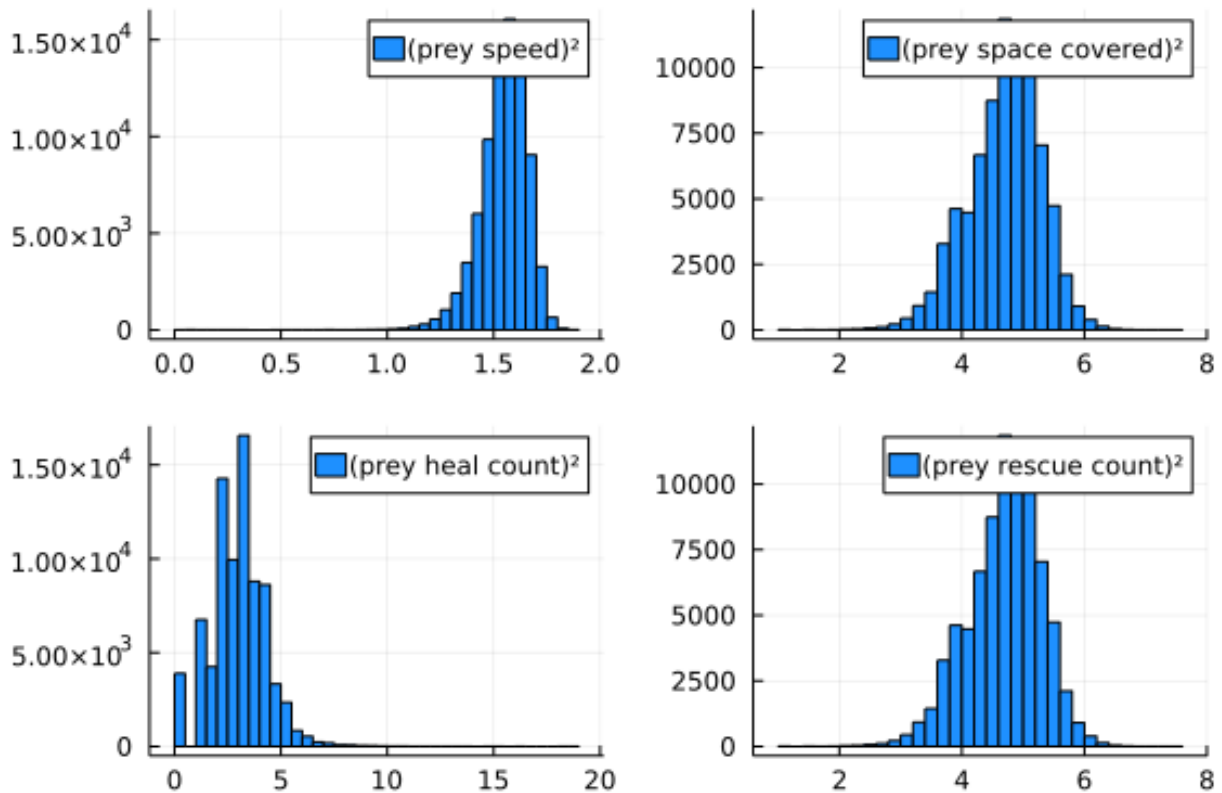
```
f = histogram(data.prey_avg_speed,  
              bins = 40,  
              label = "prey speed",  
              color = :dodgerblue1)  
g = histogram(data.prey_avg_amount_tiles_visited,  
              bins = 40,  
              label = "prey space covered",  
              color = :dodgerblue1)  
h = histogram(data.prey_total_heal_count,  
              bins = 40,  
              label = "prey heal count",  
              color = :dodgerblue1)  
i = histogram(data.prey_avg_amount_tiles_visited,  
              bins = 40,  
              label = "prey rescue count",  
              color = :dodgerblue1)  
  
plot(f, g, h, i, layout = 4)
```



2.2.2 Square root-transformed variables

```
f1 = histogram(sqrt.(data.prey_avg_speed),
               bins = 40,
               label = "(prey speed)2",
               color = :dodgerblue1)
g1 = histogram(sqrt.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "(prey space covered)2",
               color = :dodgerblue1)
h1 = histogram(sqrt.(data.prey_total_heal_count),
               bins = 40,
               label = "(prey heal count)2",
               color = :dodgerblue1)
i1 = histogram(sqrt.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "(prey rescue count)2",
               color = :dodgerblue1)

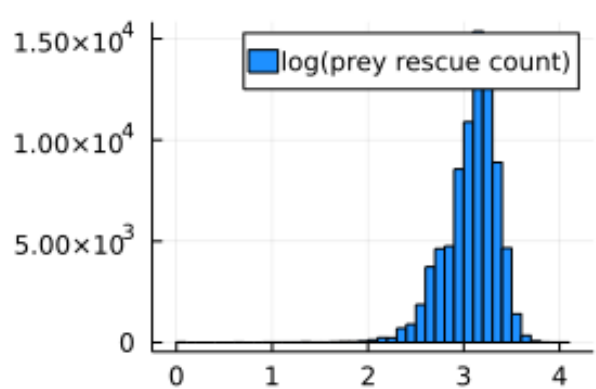
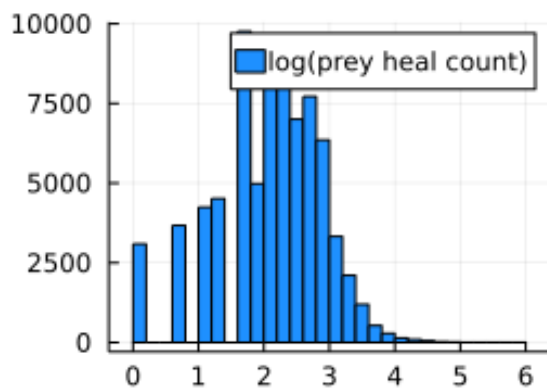
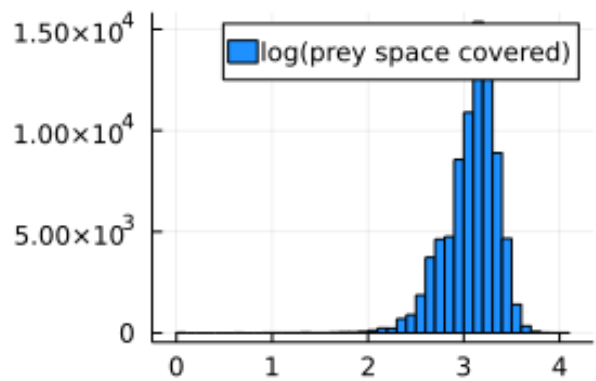
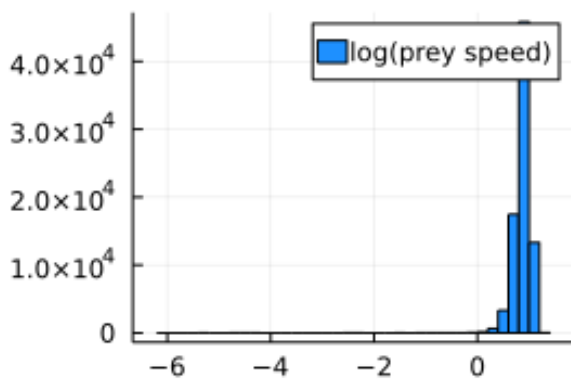
plot(f1, g1, h1, i1, layout = 4)
```



2.2.3 Log-transformed variables

```
f2 = histogram(log.(data.prey_avg_speed),
               bins = 40,
               label = "log(prexy speed)",
               color = :dodgerblue1)
g2 = histogram(log.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "log(prexy space covered)",
               color = :dodgerblue1)
h2 = histogram(log.(data.prey_total_heal_count),
               bins = 40,
               label = "log(prexy heal count)",
               color = :dodgerblue1)
i2 = histogram(log.(data.prey_avg_amount_tiles_visited),
               bins = 40,
               label = "log(prexy rescue count)",
               color = :dodgerblue1)

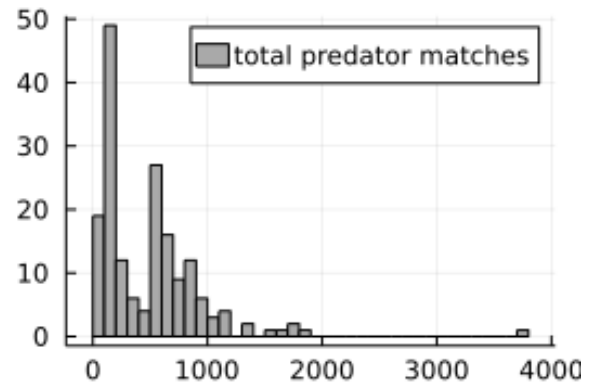
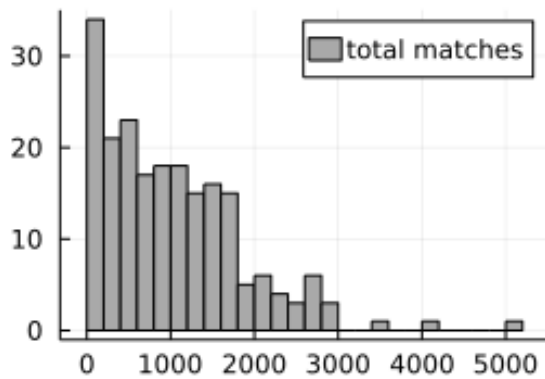
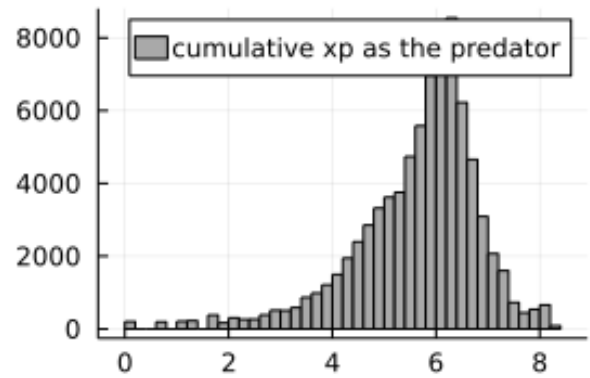
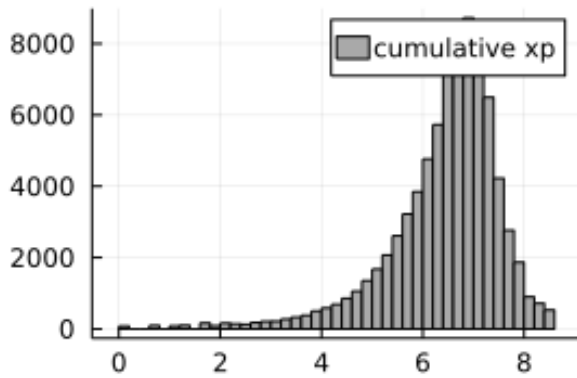
plot(f2, g2, h2, i2, layout = 4)
```



2.3 Distribution of player experience

We now investigate the distribution of the cumulative and total player experience.

```
j = histogram(log.(data.cumul_xp_total),  
              bins = 40,  
              label = "cumulative xp",  
              color = :darkgray)  
k = histogram(log.(data.cumul_xp_role),  
              bins = 40,  
              label = "cumulative xp as the predator",  
              color = :darkgray)  
l = histogram(unique(data.total_matches),  
              bins = 40,  
              label = "total matches",  
              color = :darkgray)  
m = histogram(unique(data.total_matches_killer),  
              bins = 40,  
              label = "total predator matches",  
              color = :darkgray)  
  
plot(j, k, l, m, layout = 4)
```



3. Transformed-variables selected for the analyses