Data exploration

Anonymous for review

2021-05-20

1. Prepare the data

1.1 Load the required packages

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

1.2 Import the data

```
data = DataFrame(CSV.File("data/merged-data.csv"))
```

2. Inspect the data

2.1 Distribution of predator behaviour

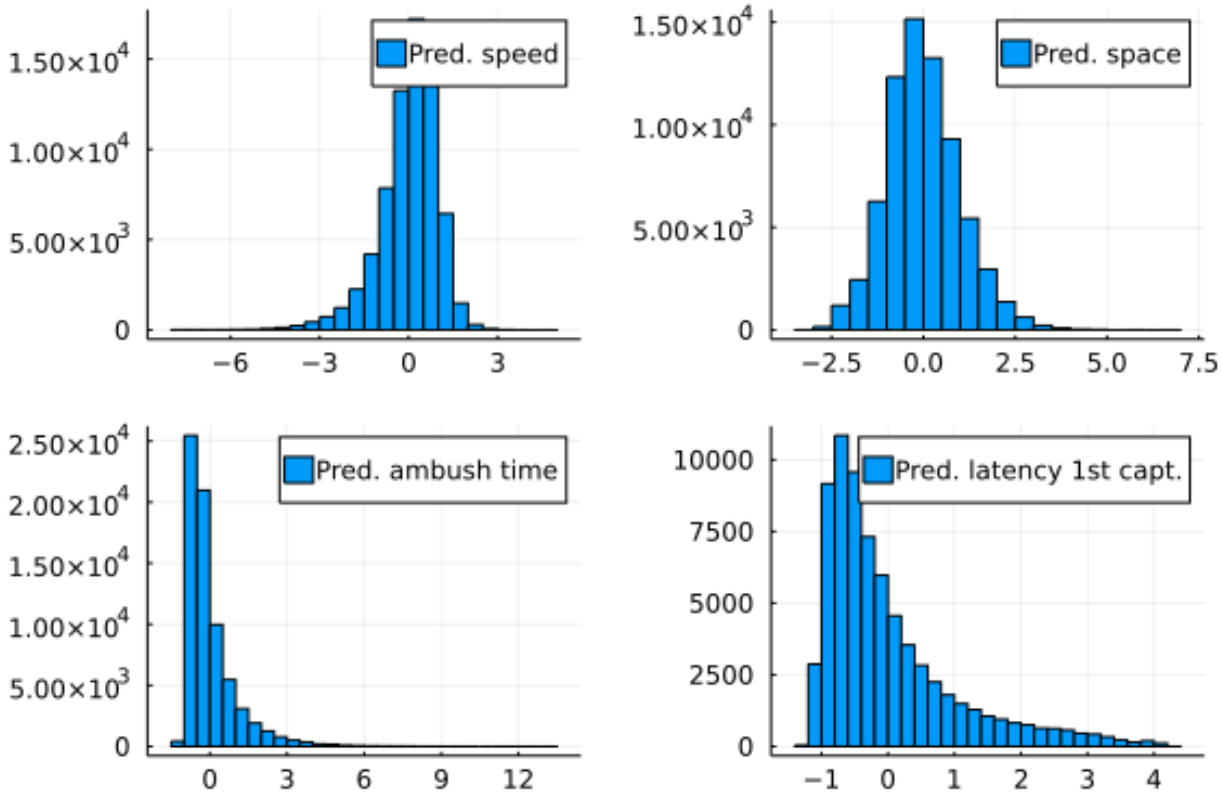
In this section, we will inspect the distribution of the different predator behaviors. We provide the distribution of the raw variables and their transformations.

The transformations include :

```
- square root
- log
```

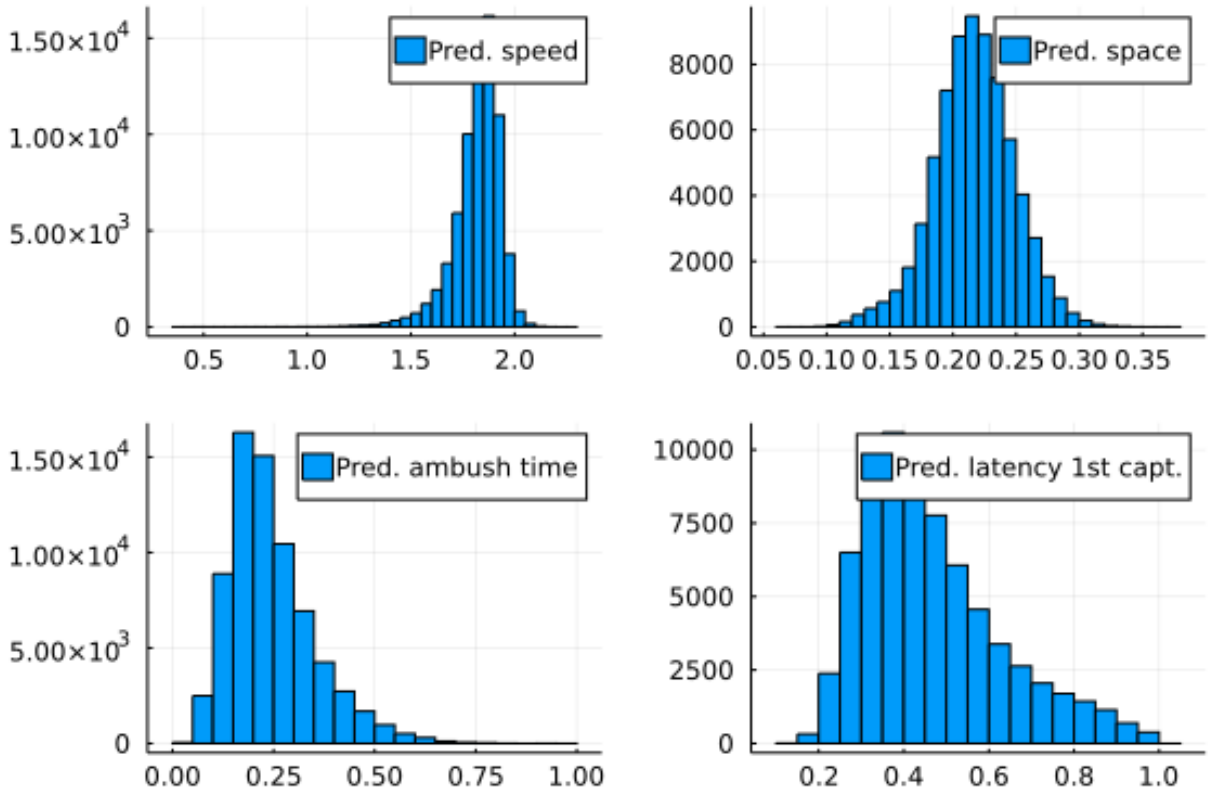
2.1.1 Raw variables

```
# Predator behavior
a = histogram(data.Zspeed, bins = 40,
                label = "Pred. speed")
b = histogram(data.Zspace_covered_rate, bins = 40,
                label = "Pred. space")
c = histogram(data.Zprox_mid_guard, bins = 40,
                label = "Pred. ambush time")
d = histogram(data.Zhook_start_time, bins = 40,
                label = "Pred. latency 1st capt.")

plot(a, b, c, d, layout = 4)
```



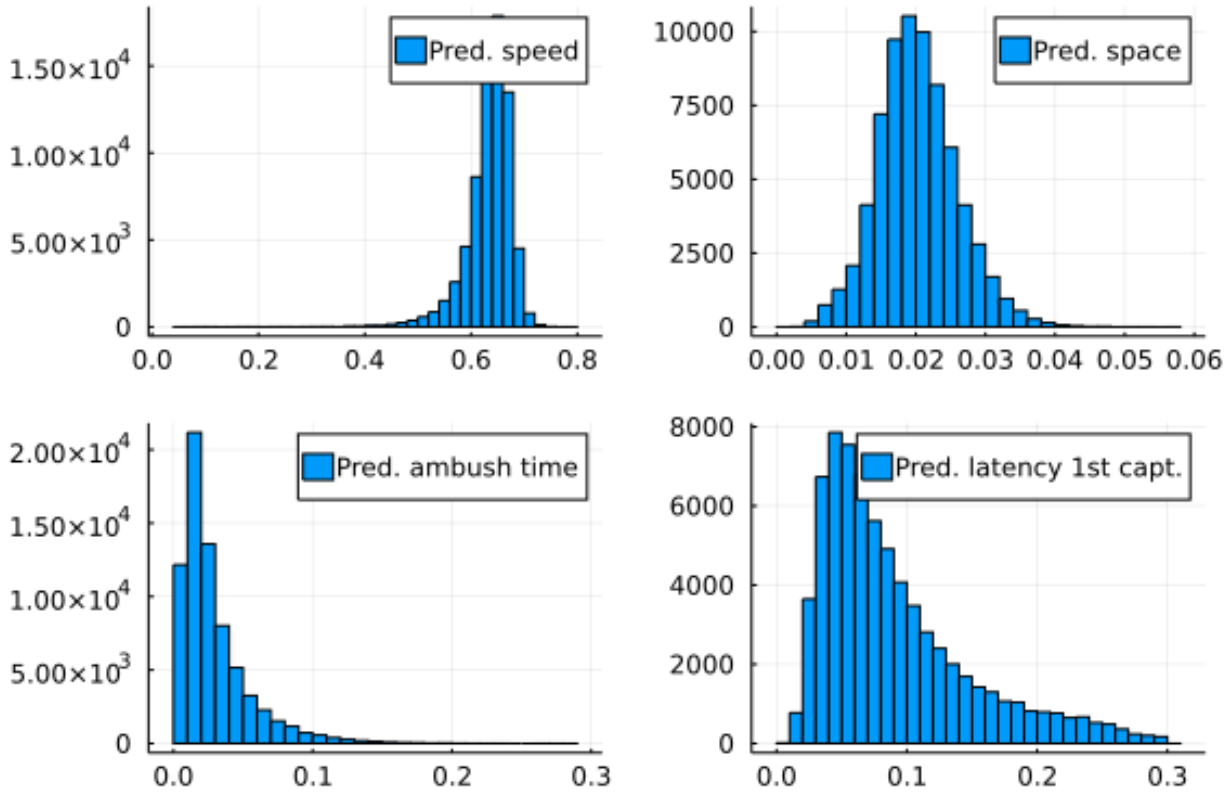
2.1.2 Square root variables

```
# Square root of predator behavior
a1 = histogram(data.sqrtspeed, bins = 40,
               label = "Pred. speed")
b1 = histogram(data.sqrtspace_covered_rate, bins = 40,
               label = "Pred. space")
c1 = histogram(data.sqrtprox_mid_guard, bins = 40,
               label = "Pred. ambush time")
d1 = histogram(data.sqrthook_start_time, bins = 40,
               label = "Pred. latency 1st capt.")

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



2.1.3 Log variables

```
# Log of predator behavior
a2 = histogram(data.logspeed, bins = 40,
               label = "Pred. speed")
b2 = histogram(data.logspace_covered_rate, bins = 40,
               label = "Pred. space")
c2 = histogram(data.logprox_mid_guard, bins = 40,
               label = "Pred. ambush time")
d2 = histogram(data.loghook_start_time, bins = 40,
               label = "Pred. latency 1st capt.")

plot(a2, b2, c2, d2, layout = 4)
```



2.1 Distribution of prey behavior

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

The transformations include :

- square root
- log

```
# Prey behavior
e = histogram(data.surv_speed, bins = 40,
                label = "Prey speed")
f = histogram(data.surv_space_covered_rate, bins = 40,
                label = "Prey space")
#g = histogram(data.boldness, bins = 50, label = "Prey boldness")
plot(e, f, layout = 2, size = (600, 200))
```