Data Exploration

Nathan Shepherd

2022-03-12

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
random_ep_observation_rewards <- read_csv("../utils/random_ep_observation_rewards.csv",show_col_types =
```

```
## New names:
## * `` -> ...1
```

```
names(random_ep_observation_rewards)
```

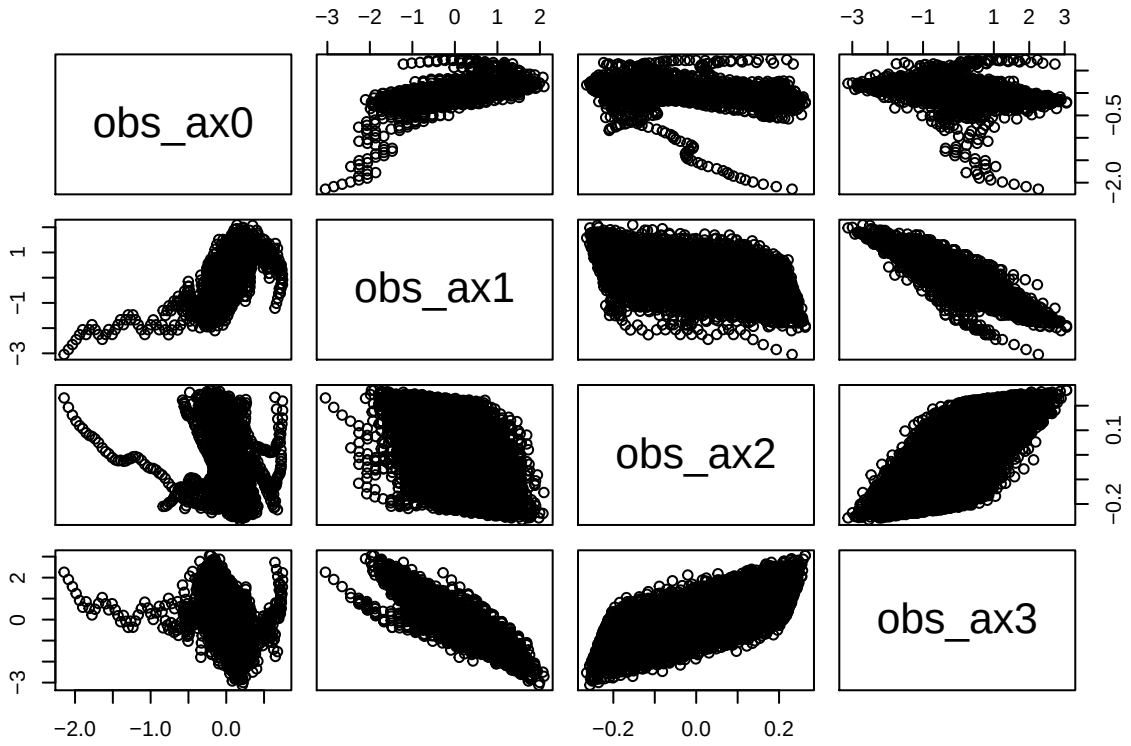
```
## [1] "...1"     "timestep" "episode"  "reward"   "obs_ax0"  "obs_ax1"  "obs_ax2"
## [8] "obs_ax3"
```

```
rand_dat = random_ep_observation_rewards[2:8] # drop index column
```

```
summary(rand_dat)
```

```
##     timestep         episode          reward          obs_ax0
##  Min.   :  0.00   Min.   :  0.0   Min.   :  8.00   Min.   :-2.1444235
##  1st Qu.:  5.00   1st Qu.:253.0   1st Qu.: 17.00   1st Qu.:-0.0413438
##  Median : 11.00   Median :497.0   Median : 23.00   Median : 0.0018257
##  Mean   : 13.96   Mean   :498.2   Mean   : 27.91   Mean   : 0.0008848
##  3rd Qu.: 19.00   3rd Qu.:747.0   3rd Qu.: 35.00   3rd Qu.: 0.0450811
##  Max.   :101.00   Max.   :999.0   Max.   :101.00   Max.   : 0.7452676
##     obs_ax1             obs_ax2             obs_ax3
##  Min.   :-3.045133   Min.   :-0.2641969   Min.   :-3.1194220
##  1st Qu.:-0.376045   1st Qu.:-0.0556148   1st Qu.:-0.5547729
##  Median :-0.002044   Median :-0.0008753   Median : 0.0009272
##  Mean   :-0.007947   Mean   :-0.0007722   Mean   : 0.0040934
##  3rd Qu.: 0.369290   3rd Qu.: 0.0539384   3rd Qu.: 0.5556752
##  Max.   : 2.094117   Max.   : 0.2631265   Max.   : 3.0541437
```

```
pairs(rand_dat[4:7])
```




```
by_timestep = rand_dat %>% group_by(timestep)
timestep_summ = summarize(by_timestep, count=n(), avg.reward=mean(reward))
ggplot(timestep_summ, aes(x=timestep, y=count)) + geom_point()
```