Data Exploration

Nathan Shepherd

2022-03-24

```
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)
```

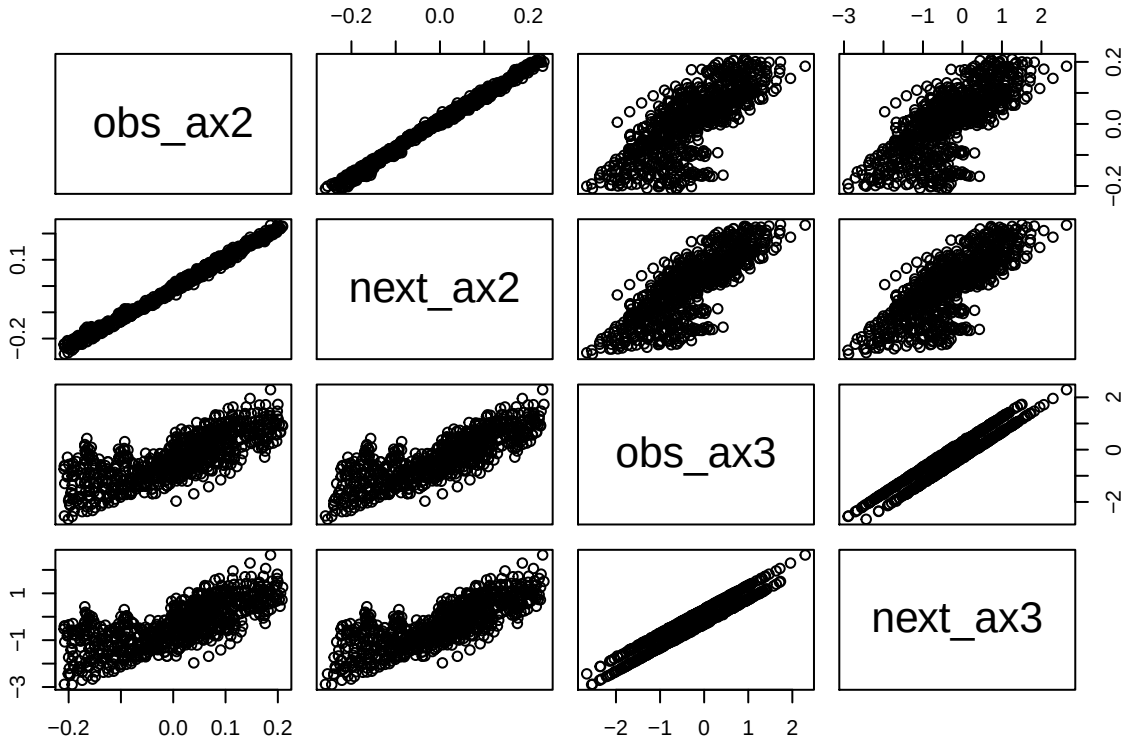
```
rand_dat <- read_csv("../utils/rand_state_acts.csv",show_col_types = FALSE)
names(rand_dat)
```

```
##  [1] "timestep" "episode"  "reward"   "act"      "obs_ax0"  "next_ax0"
##  [7] "obs_ax1"  "next_ax1" "obs_ax2"  "next_ax2" "obs_ax3"  "next_ax3"
```

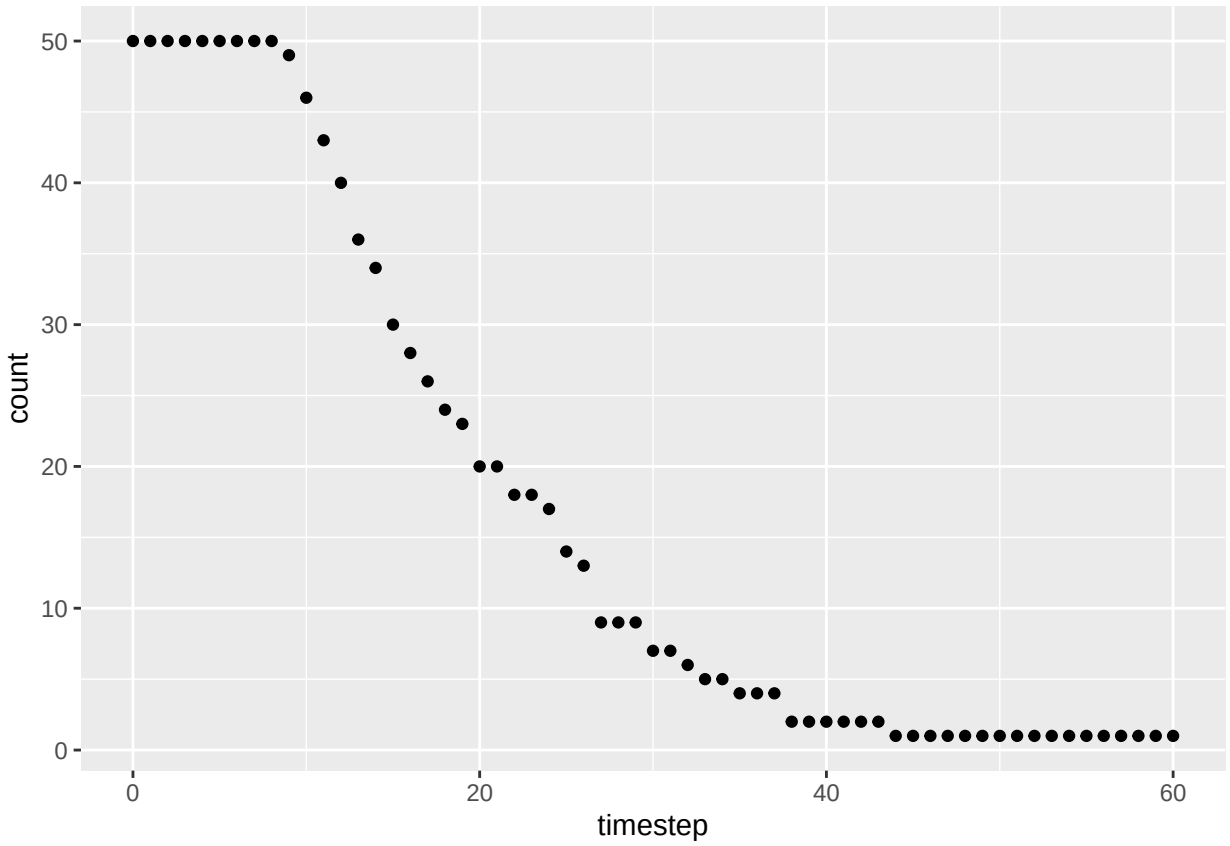
```
summary(rand_dat)
```

```
##     timestep        episode          reward           act
##  Min.   : 0.00   Min.   : 0.00   Min.   : 9.00   Min.   :0.0000
##  1st Qu.: 5.00   1st Qu.:14.00   1st Qu.:16.00   1st Qu.:0.0000
##  Median :10.00   Median :26.00   Median :25.00   Median :1.0000
##  Mean   :12.46   Mean   :25.53   Mean   :25.92   Mean   :0.5568
##  3rd Qu.:18.00   3rd Qu.:36.00   3rd Qu.:32.00   3rd Qu.:1.0000
##  Max.   :60.00   Max.   :49.00   Max.   :61.00   Max.   :1.0000
##     obs_ax0             next_ax0          obs_ax1            next_ax1
##  Min.   :-0.1096541   Min.   :-0.12996   Min.   :-1.40519   Min.   :-1.6015
##  1st Qu.:-0.0156297   1st Qu.:-0.01832   1st Qu.:-0.06116   1st Qu.:-0.2266
##  Median :-0.0000219   Median : 0.00432   Median : 0.16127   Median : 0.1625
##  Mean   : 0.0242654   Mean   : 0.02875   Mean   : 0.22419   Mean   : 0.2461
##  3rd Qu.: 0.0597766   3rd Qu.: 0.07109   3rd Qu.: 0.55266   3rd Qu.: 0.5563
##  Max.   : 0.5181503   Max.   : 0.55580   Max.   : 1.88238   Max.   : 2.0789
##     obs_ax2            next_ax2           obs_ax3           next_ax3
##  Min.   :-0.20863   Min.   :-0.25873   Min.   :-2.6632   Min.   :-2.8973
##  1st Qu.:-0.03345   1st Qu.:-0.05214   1st Qu.:-0.7747   1st Qu.:-0.8118
##  Median : 0.03682   Median : 0.03486   Median :-0.1940   Median :-0.2275
##  Mean   : 0.01503   Mean   : 0.01036   Mean   :-0.2336   Mean   :-0.2620
##  3rd Qu.: 0.06781   3rd Qu.: 0.07017   3rd Qu.: 0.3324   3rd Qu.: 0.3572
##  Max.   : 0.20913   Max.   : 0.23477   Max.   : 2.2920   Max.   : 2.6358
```

```
pairs(rand_dat[9:12])
```




```
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()
```


50
40
30
20
10
0
count
0
20
40
60
timestep