The Great Penguin Stretching Project

Bio 359K - Compiled by May Dixon and Caitlin Leslie

4/20/2020

Hypothesis:

Penguins, being social birds, do contagious stretching. They are more likely to stretch when a nearby bird has just stretched.

This document includes:

- Summary stats
  - mean median interstretch interval
  - mean and median for penguin group size
- Frequencies of inter-stretch intervals
  - Full histogram of interstretch intervals for all data
  - Histogram of interstretch intervals seperated by group size
  - Shapiro- wilks and KS tests of normality
- Exploration of group size and inter-stretch interval
  - Plot of the average interstretch interval per group-size
  - Plot of the average interstretch interval per group-size per individual(*group size)
  - regressions of slope and fit for these data
- OTHER interesting data visualized:
  - When did people observe?
  - Did average stretch rate vary by time of day?
  - Did group size vary by time of day?

First, What was the mean and median inter-stretch interval compiled between all group sizes?

Mean interstretch interval:

[1] 132.2264

Median interstretch interval:

[1] 64

Prediction #1:

If penguins have contagious behaviors, then we expect there to be bouts of stretching that are clustered, wherein there are short bursts of stretching caused by a priming individual, followed by longer periods of no stretching when no individual has stimulated stretching. Numerically, this means we expect there to be an over-abundance of very short inter-stretch intervals and very long inter-stretch intervals, relative to random expectations.

Histograms

Histogram 1: Full data, cut off at 1 hr, 10 s bins.
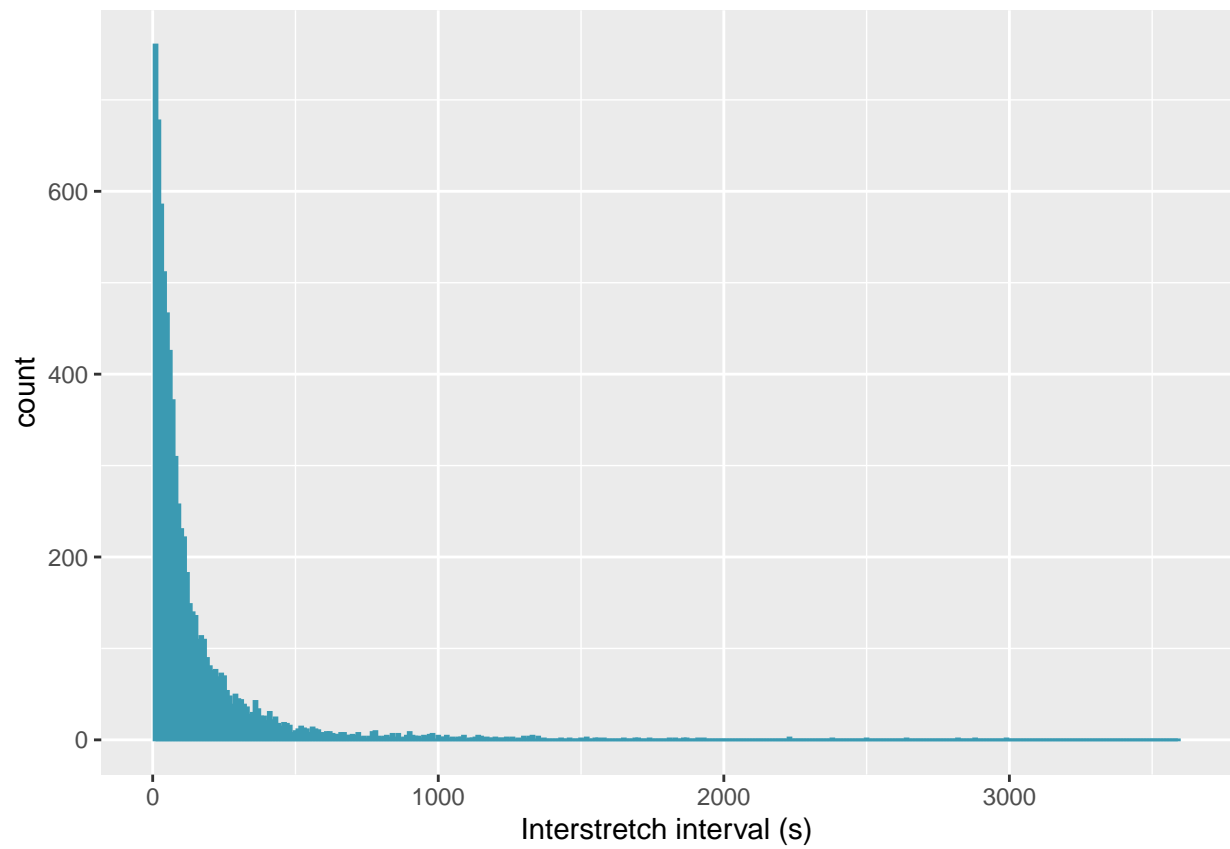
Histogram 2: Same data, zoomed in to a maximum of 300 s, 10 s bins.
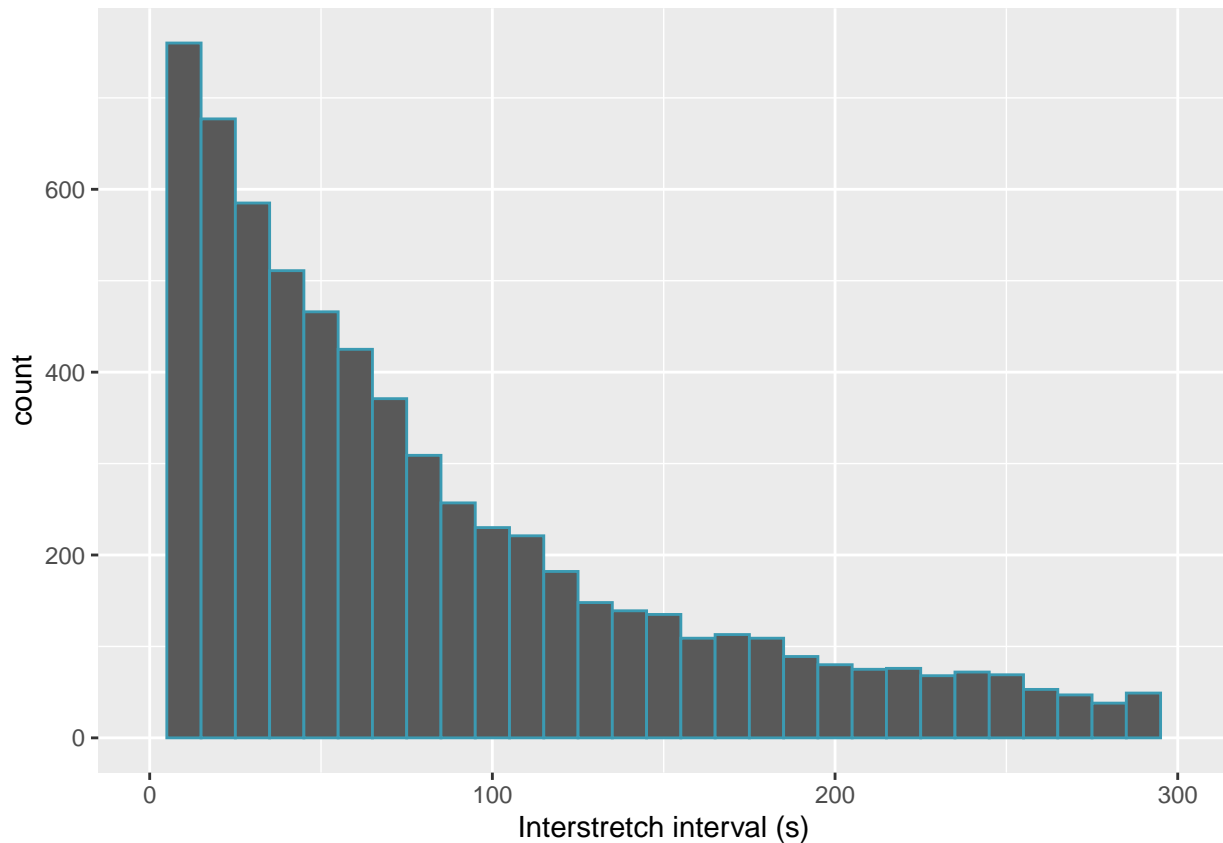
Even just visually, this is clearly a very non-normal distribution. Rather, this is a poisson distribution, with an over-abundance of low-value inter-stretch intervals. This is consistent with our expectations. Let's test this statistically:

Stats: is this distribution normal?

KS test and Shapiro test of normality:

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  penguin_subdata$Interstretch_interval_s and rpois(5000, 10)
## D = 0.8058, p-value < 2.2e-16
## alternative hypothesis: two-sided

##
##  Shapiro-Wilk normality test
##
## data:  penguin_subdata$Interstretch_interval_s
## W = 0.058706, p-value < 2.2e-16
```

No, this distribution is decidedly non- normal.

But is it not normal at each group size? It is possible that each group has its own inter-stretch interval, and combining all the group sizes makes the distribution appear non-normal when it really is. We can check this by seperating out the histograms by group size:

Histogram #3: Interstretch interval broken down by group size (bins of 10s)
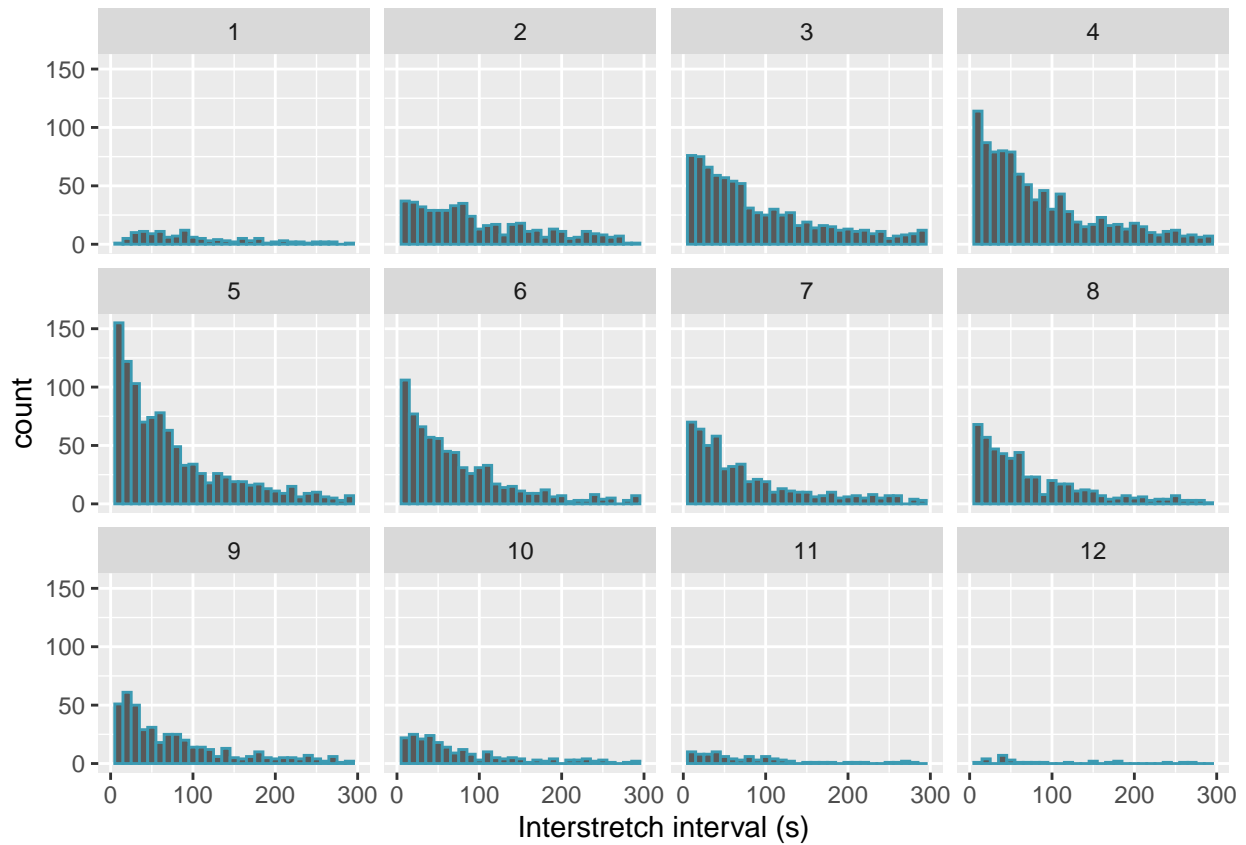
- Immediately, we can see that some of the group sizes have a lot more data than others
- For most of the group sizes, the distribution still does not appear not normal.

Let's look closer at the distribution of stretches for groups of one and two penguins:

Distribution of stretching rates for groups of one and two penguins
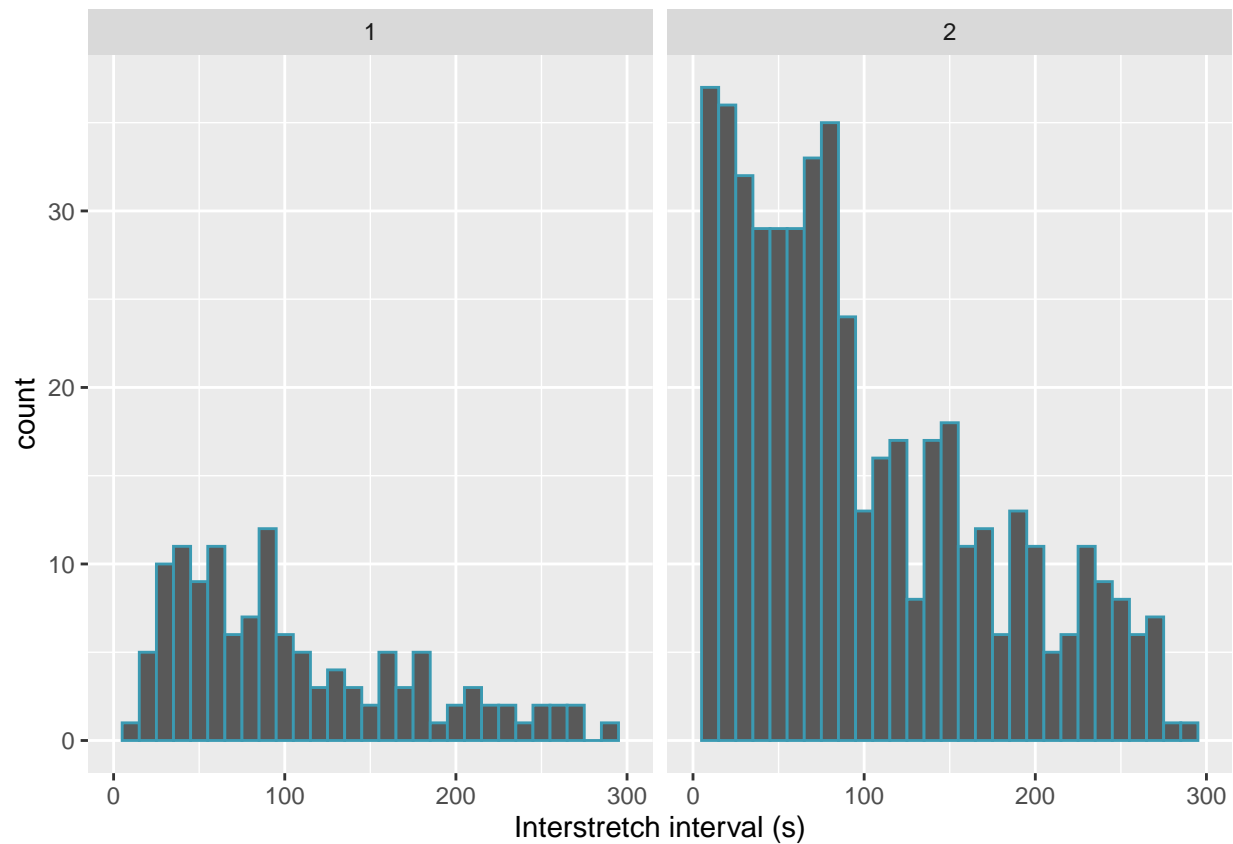
- It is interesting that observed groups of one still don't really have a normal distribution of stretching, like we might expect

- However, the group of 1 is the most normal of all the groups, and is the only group where there are more observations in the 30-40 s bin than the >20 s bin.

- Perhaps sometimes when we were observing 1 penguin, that penguin was not truly alone, and could have observed other individuals stretching

We can also formally test statistically whether the inter-stretch intervals are non-normal at each group size:

KS test at each group size

- Table 1 : The KS tests show that all the group sizes were non-normal, including the observations with only 1 penguin in a group

Prediction #2: Group size

- Another prediction is that there will be more stretches per individual in larger groups than in smaller groups, because any individual is more likely to be stimulated by another individual in the group. Numerically, this means we expect a positive correlation between group size and # of stretches per individual.

- We might also expect there the be a shorter inter-stretch interval per penguin for larger group sizes

Table 1: Summary of normality tests

n_penguins	shapiro_statistic	shapiro_p.value	KStest_statistic	KStest_p.value
1	0.6281271	0.0000000	0.9878070	0
2	0.6360049	0.0000000	0.9697126	0
3	0.6200756	0.0000000	0.9464000	0
4	0.5479714	0.0000000	0.9339000	0
5	0.5141858	0.0000000	0.8950926	0
6	0.5009664	0.0000000	0.8894247	0
7	0.6694460	0.0000000	0.8864563	0
8	0.4202551	0.0000000	0.9089182	0
9	0.6487368	0.0000000	0.9190424	0
10	0.6259539	0.0000000	0.9366752	0
11	0.7509499	0.0000000	0.9184750	0
12	0.7787111	0.0002366	1.0000000	0

Table 2: Stretching by group size

| n_penguins	stretch_mean	stretch_sd	stretch_median	n_obs	stretch_mean_penguin	stretch_med_penguin
1	272.70500	361.4757	162.0	200	272.7050	162.0
2	198.54874	258.3485	108.0	636	397.0975	216.0
3	158.04665	226.8575	77.5	986	474.1400	232.5
4	136.23652	214.5862	68.0	1150	544.9461	272.0
5	107.95697	198.8644	54.0	1255	539.7849	270.0
6	123.49203	247.7682	53.0	878	740.9522	318.0
7	111.83460	165.8954	50.0	659	782.8422	350.0
8	101.48911	172.6994	54.0	597	811.9129	432.0
9	92.78958	118.8590	53.0	499	835.1062	477.0
10	98.40244	135.1804	51.5	246	984.0244	515.0
11	101.95918	114.1070	59.5	98	1121.5510	654.5
12	147.05556	141.9295	77.5	36	1764.6667	930.0

How did average inter-stretch interval vary by group size?

Here, we can summarize the mean and median inter-stretch intervals observed at each group size, and also control for the number of penguins present by multiplying the inter-stretch interval by the # of penguins in the group (Table 2)

- n_penguins = group size
- stretch_mean = interstretch interval mean
- stretch_sd = interstretch interval standard deviation
- stretch_median = median interstretch interval
- n_obs = number of observations / group
- stretch_mean_penguin = mean interstretch interval / # penguins in group
- stretch_med_penguin = mean interstretch interval / # penguins in group

Let's plot one of these variables. We can see how inter-stretch interval varies with group size, for all the penguins in a group:

Inter-stretch interval by group size:
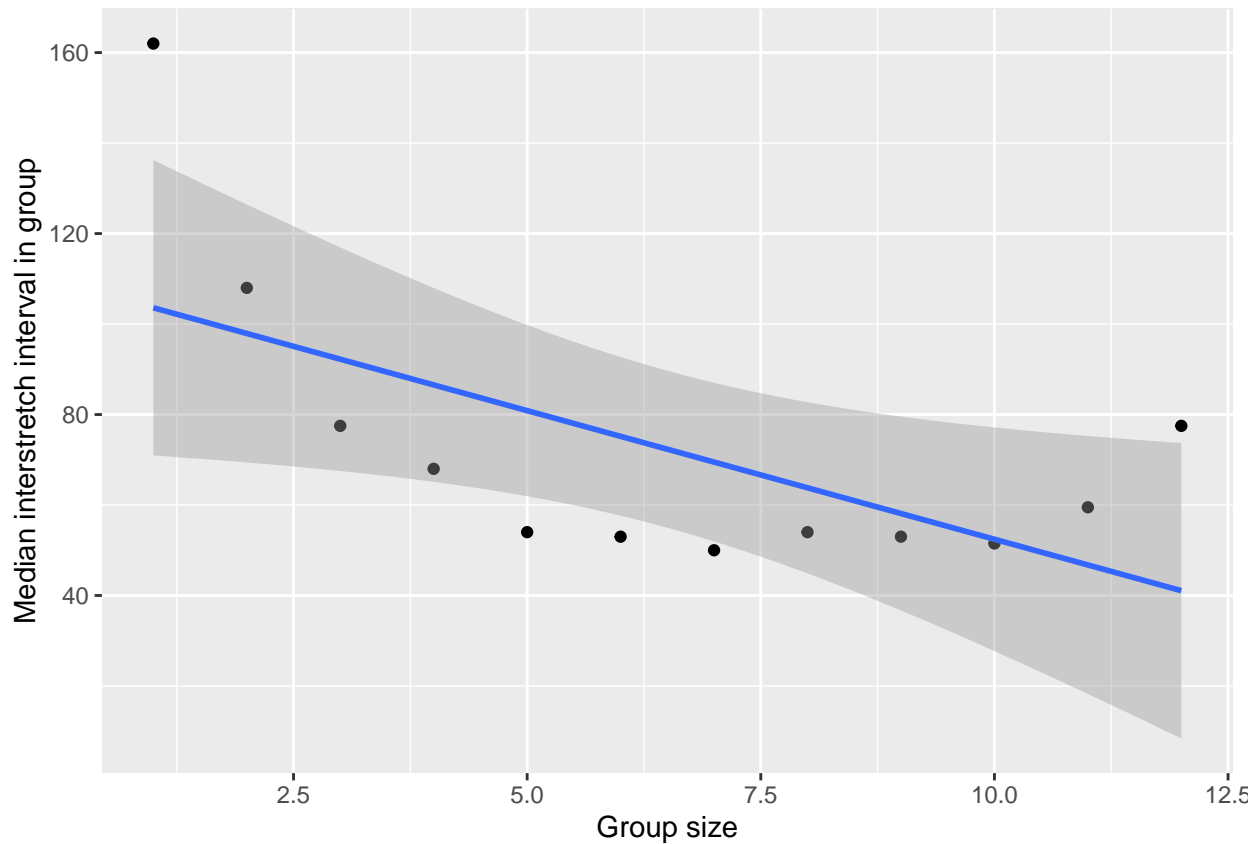
- The data look a bit parabolic, but in general there looks to be a negative correlation, as group size gets bigger, time between stretches goes down.

Let's test this formally with a linear regression:
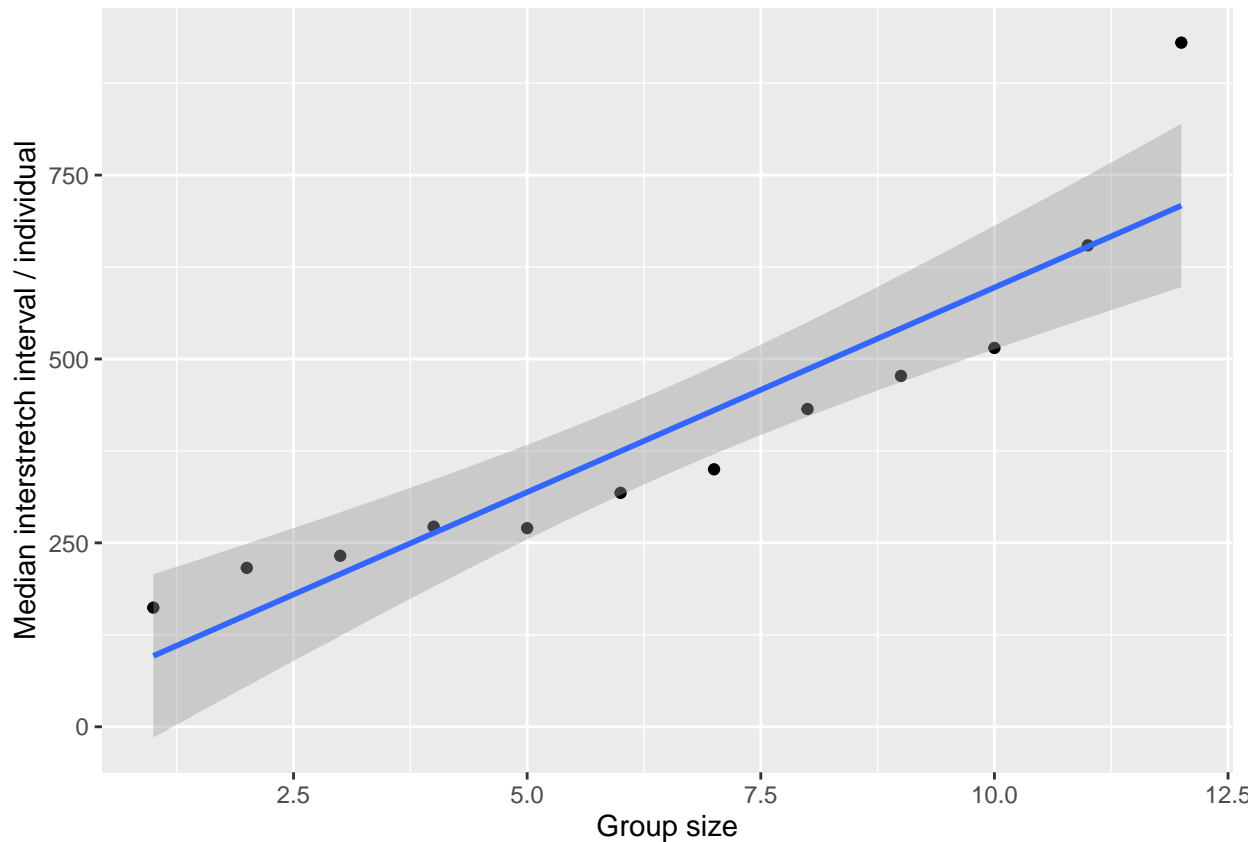
Linear Regression

```
##
## Call:
## lm(formula = penguin_group$inter_stretch_intervals_s ~ penguin_group$n_penguins)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -185.8 -105.6  -62.0   21.3 3783.4
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               198.287      6.080   32.61   <2e-16 ***
## penguin_group$n_penguins  -12.448      1.044  -11.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 213.1 on 7238 degrees of freedom
## Multiple R-squared:  0.01926,    Adjusted R-squared:  0.01913
## F-statistic: 142.2 on 1 and 7238 DF,  p-value: < 2.2e-16
```

- Basically, the inter-stretch interval decreases by about 12 seconds for every increase of 1 penguin in a

group

- BUT of course, if you have more penguins around, the interstretch interval will go down even if each penguin is just stretching at the same rate. So we should also account for the # of individuals around by multiplying the interstretch interval by the # of penguins around.

Interstretch interval per individual group size/ individual

- Now that we have controlled for the # of penguins,we can see that the interstretch interval is closely correlated with group size, and there is a positive correlation, where the stretch rate/ individual increases with group size!

- We can test this formally:

Regression by group size

```
##
## Call:
## lm(formula = penguin_group$is_p_group ~ penguin_group$n_penguins)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1081.1  -485.0  -276.7    93.4 22759.0
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               224.967     30.003   7.498 7.25e-14 ***
## penguin_group$n_penguins   76.343      5.152  14.819  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1051 on 7238 degrees of freedom
## Multiple R-squared:  0.02945,    Adjusted R-squared:  0.02931
## F-statistic: 219.6 on 1 and 7238 DF,  p-value: < 2.2e-16
```

- So basically, individual penguins stretch less often in larger groups. This is opposite to our predictions! Why might this be?

    1. Perhaps stretching is not contagious after all
    2. Perhaps larger groups occur when penguins are being more active, and have less internal need to stretch
    3. *Perhaps penguins DO stretch socially, but it is nto dependent on group size: Pehaps contagious stretching only works short range, and it does not extend across a big group.*

I also just want to point out how skewed out data are (as is obvious in our histograms). This plot shows the full data of interstretch intervals/ individual, in black and the pink diamonds show the median values per each group size:

Interstretch interval per individual by group size, full dataset
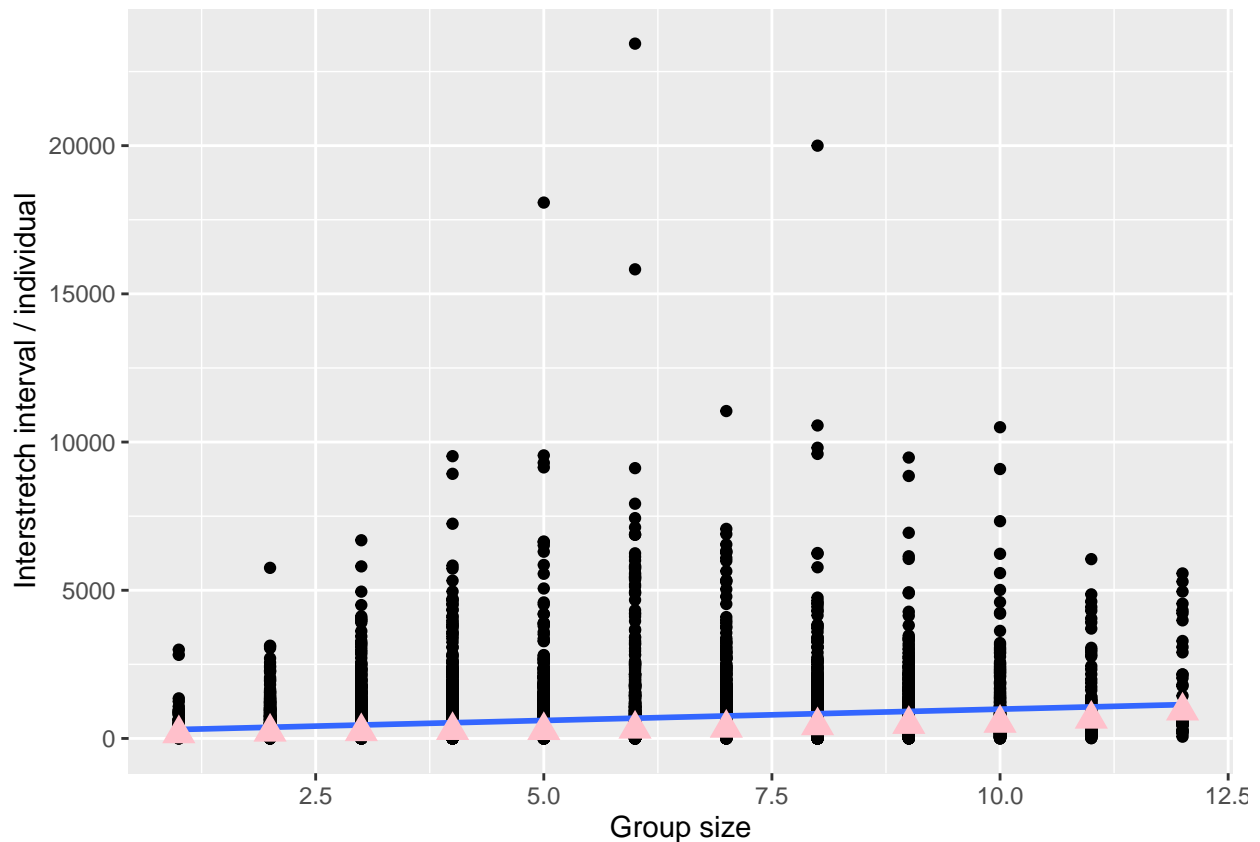
The trend is the same, as group size gets bigger, the time between stretches (per individual) increases by 75s / group member. (But now you can also see all the variance in the data. Such is the story with animal behavior! Nature is messy)

Prediction #3: There should be more observations of stretching in higher group sizes since more penguins should be influenced to stretch by a single stretchers in bigger groups

This is the number of observations of stretching at each group size:
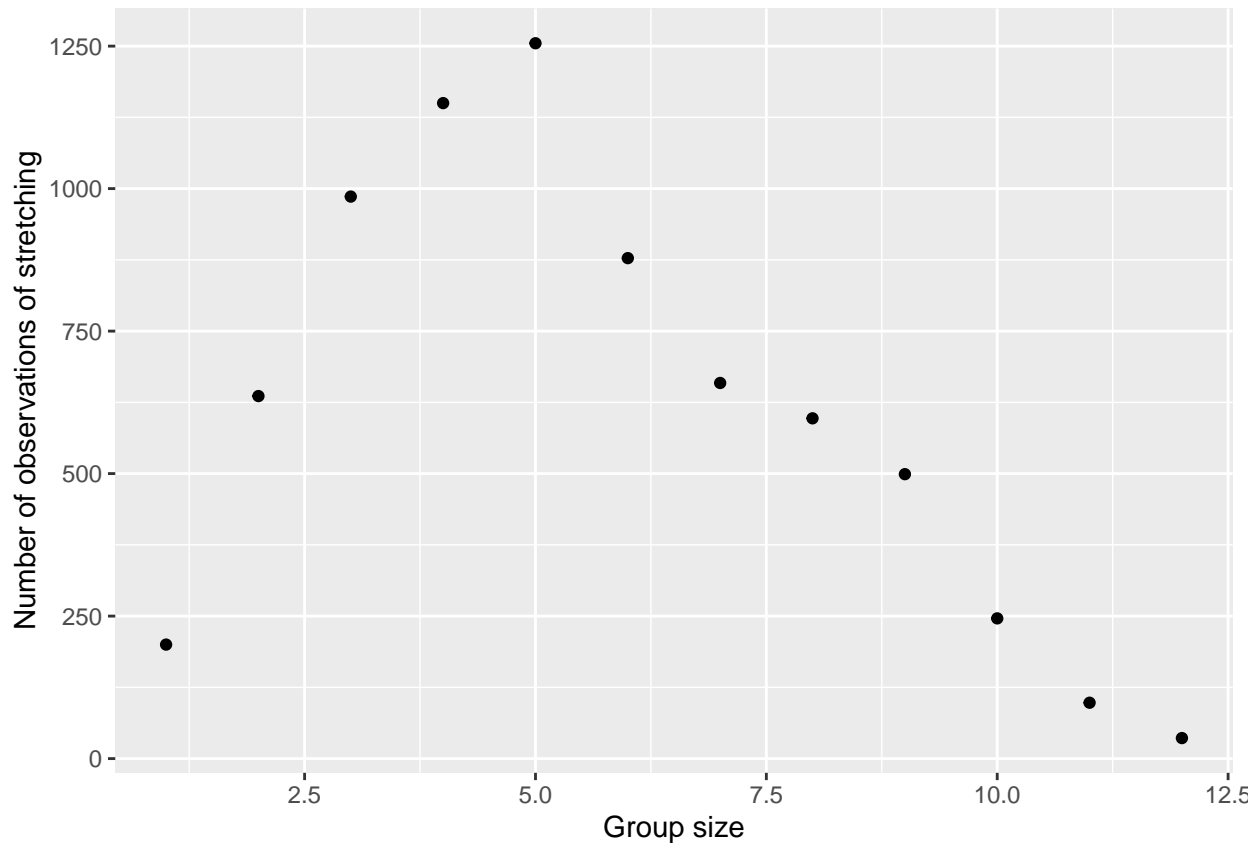
- It turns out that we observed the different groups different amounts!! So, what we are actually seeing is that we were most likely to observe groups of 5 than any other group! This is why it is always important to critically examine our own assumptions. If we had examined each group size for the same amount of time, then this measure may have been more useful. Now, we simply know that the commonest group size (or the favorite # of penguins for the camera-person to observe), is 5, with a range from 0- 14 penguins.

As long as we have the data, we can also answer some other questions, just for "fun":

- What dates did you observe on?
- When did ya'll observe?
- Did median stretch interval change throughout the day? (Maybe they should stretch more in the morning?)
- Did group size vary throughout the day?

What dates did ya'll do most of your observations?

- How many observations happened on each date?
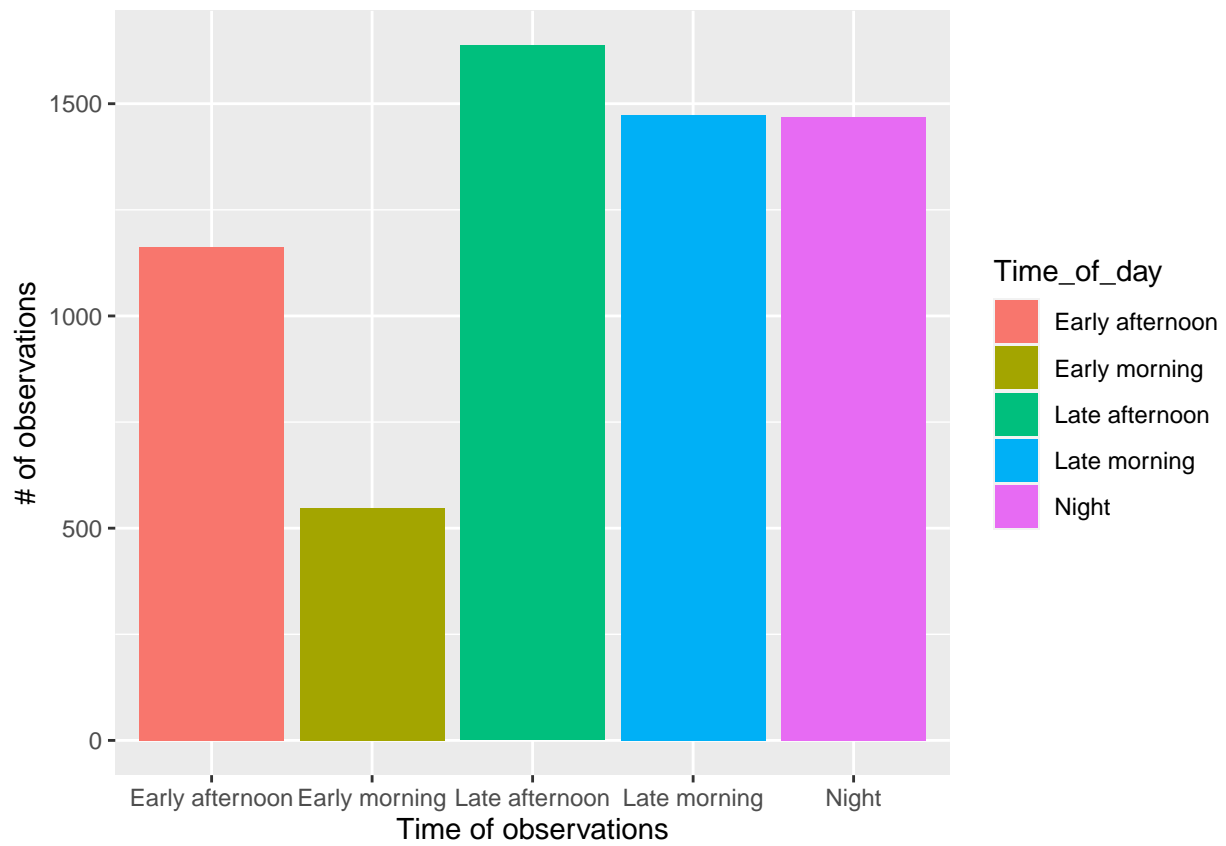
Table 3: Time of day data

Time_of_day	number_observations	median_intervals	median_groupsize
Early afternoon	1163	84	4
Early morning	548	57	7
Late afternoon	1637	69	5
Late morning	1473	52	5
Night	1468	59	5
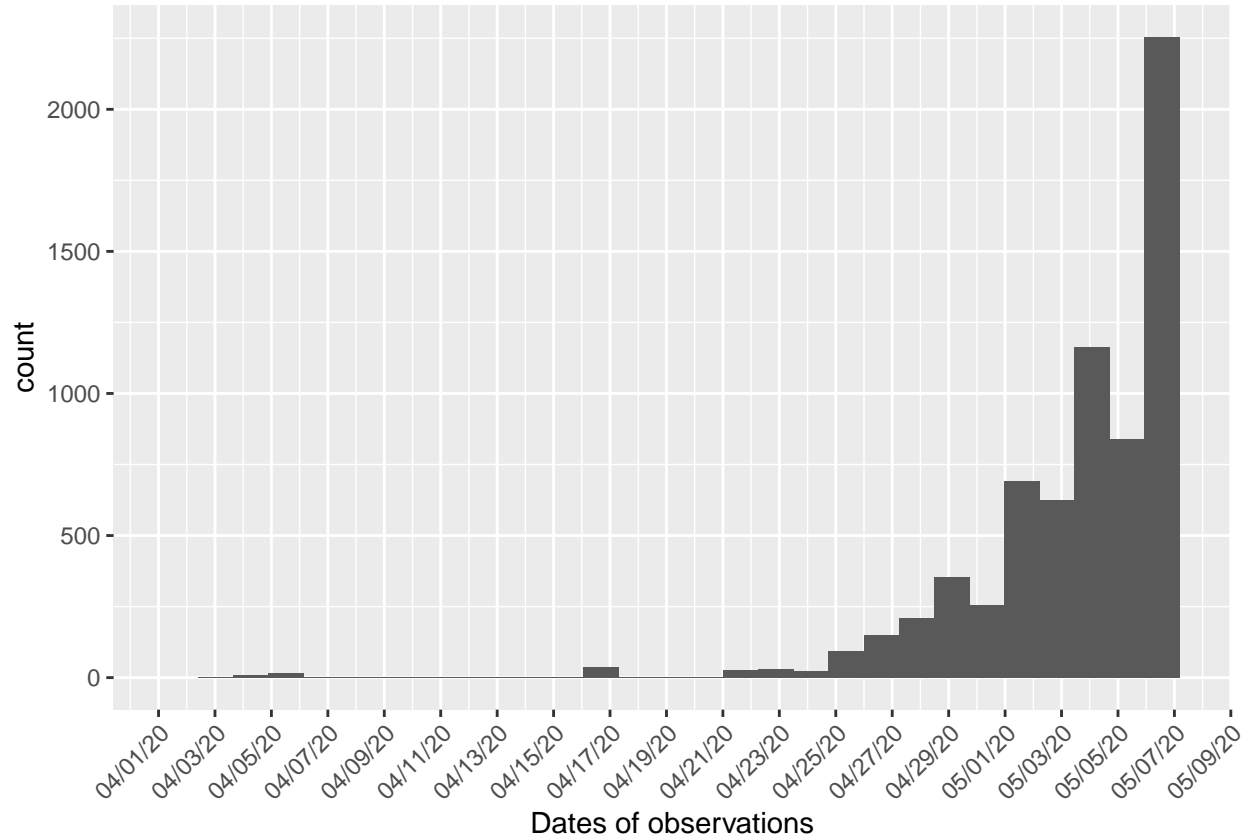
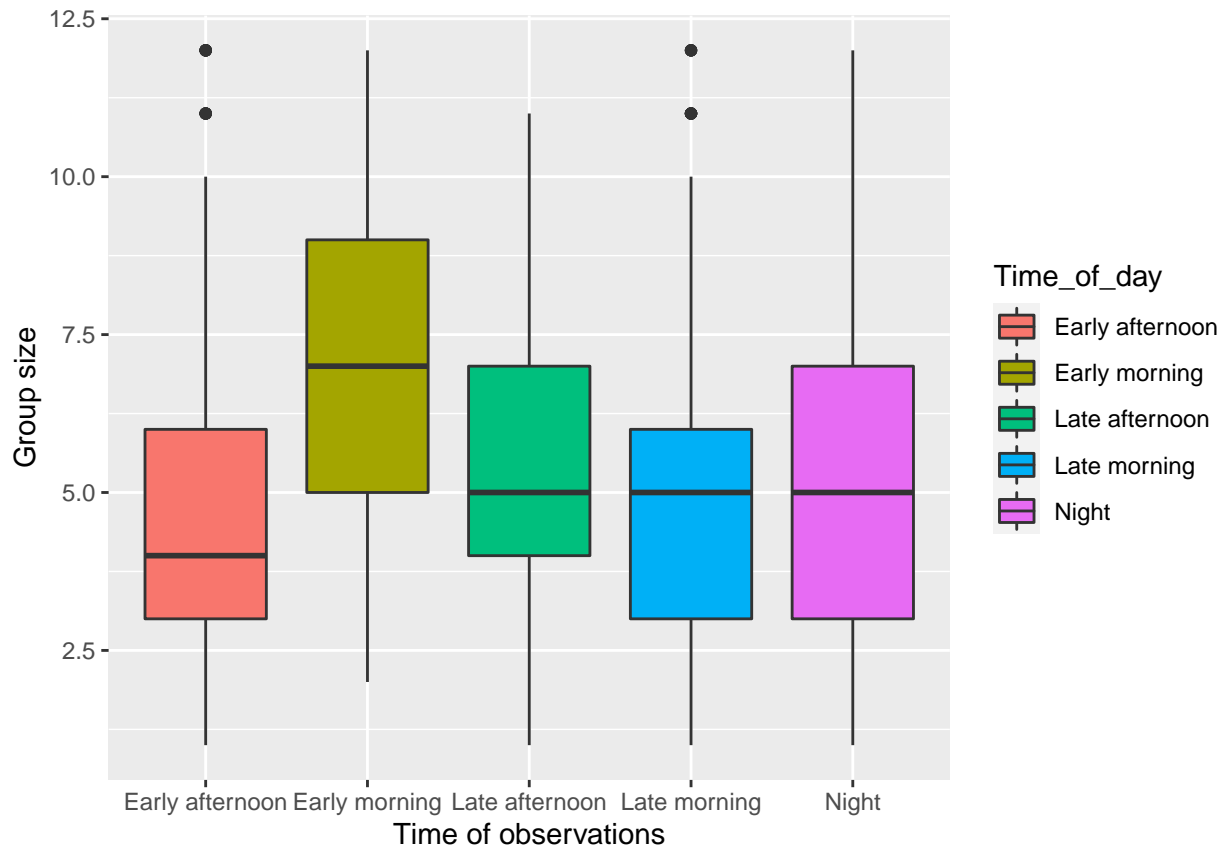Deadlines are motivating, aren't they?

Table 3: Summary stats for the time of day:

What time of day did you observe?

You were most likely to observe in early afternoon, and least likely to observe in early morning.

Did ### Did penguins hang out in different group sizes at different times of day? Plot of group size by time of day:

- Groups were perhaps slightly larger in the mornings than in the early afternoon.

Take-aways and follow ups?

Our data were mixed! It is possible that stretching is contagious, but not dependent on group size, as we predicted. Perhaps an individual penguin only pays attention to a small number of other penguins at a time, regardless of the overall group size present. This could be for social reasons, or simply a limit to whatever sensory system they use in this continuous behavior (we assumed sight, but some of you pointed out that it could be sound).

As a follow -up, we could test these same penguins slightly differently: we could see whether penguins that were closer to a stretching penguin are more likely to stretch than penguins that are farther away.