FinalReport_ReproducingPowell2017

Fiona Lightbody

12/5/2020

Final Report: Reproducing Powell et Al. 2017

Task 1)

1. The purpose of this file is to reproduce statistical figures from (Powell et al. 2017).

2. My first task is to reproduce Figure 2, featuring two boxplots of nest-entrance size from data collected in natural colonies (in the field). One plot represents the areas in mm^2, while the other is standardized against the head area of a soldier ant (2.9 mm^2).

3. The data were readily available from an associated link alongside the article, and was successfully downloaded as a CSV file. Data is available on GitHub: https://github.com/flightbody/myrepo.git

4. The data reads into R without problems. The code below was added to create a column with the standardized entrance areas, which is necessary for the figure. For this task, the readr and tidyverse libraries are recalled.

5. Metadata and code associated with this paper were not made available with its data. However, email correspondence with Dr. Powell did easily provide access. The code was initially written using a different dataset than is provided , which initially presented a challenge, but is thoroughly commented which aided in teasing out specific lines needed. Figure 2 was successfully reproduced using the original code, with the addition of the relative entrance area column, Collection_data$rel_area, as discussed in the step above. Some lines are remnants from the original code which have been commented out in the interest of precise reproduction.

```
#By Entrance Area
par(pty="s",bty="l",tcl=0.02,mgp=c(2.4,0.6,0))

boxplot(Collection_data$ent_area_mm, cex.axis=0.8, ylim=c(2,39),las=1,log="y", xant="n", xlab=expression

  #points(1,mean(Collection_data$ent_area_mm),pch=4,col="black")

ytemptck<-c(2,3,4,5,6,7,8,9,10,20,30,40)
axis(2, at=ytemptck, tck=-0.02, label=F, lwd=0, lwd.ticks=1)
axis(2, at=c(2,5,10,40), label=c(2,5,10,40), mgp=c(3,0.6,0), lwd=0, cex.axis=0.8, las=1)
```
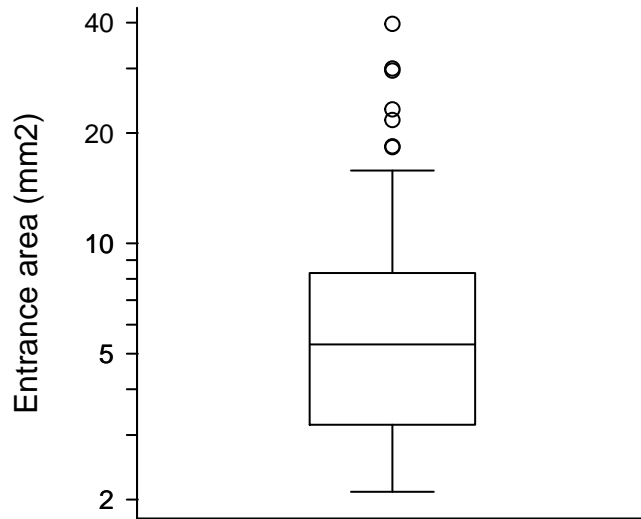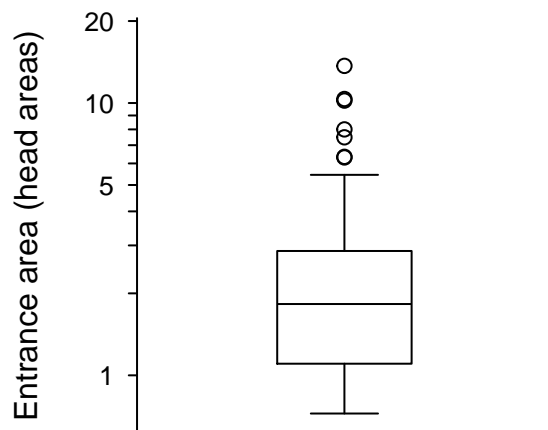
C.rohweri nests

```r
#By Head Area
par(pty="s", bty="l", tcl=-0.2, mgp=c(2.4, 0.6, 0), fin=c(4.0, 4.0))
boxplot(Collection_data$rel_area, cex.axis=0.8, ylim=c(0.7,18), log="y", xaxt="n", yaxt="n",las=1, xlab=

#points(1,mean(Collection_data$rel_area),pch=4,col="black") #To add a point for the calculated mean

ytemptck <- c(1,2,3,4,5,6,7,8,9,10,20)
axis(2, at=ytemptck, tck=-0.02, label=F, lwd=0, lwd.ticks=1)
axis(2, at=c(1,5,10,20), label=c(1,5,10,20), mgp=c(3,0.6,0), lwd=0, cex.axis=0.8, las=1)
```


C. rohweri nests

6. Overall, this task was successfully completed despite the specific datasets, for which the code was written, not being available. From the research perspective, I understand why it may be easier to clean and wrangle the dataset outside of the main code file. It leaves the two practices separate and decreases the chance of code confounding itself, but makes it much more difficult for future users who

do not understand how the code was manipulated. This presents a challenge in the next task as well. The graphs produced in this task are almost nearly identical to the published figures, except for the appearance of the outlier points. This slight difference in appearance may be due to innate differences in the data set used for the official figure compared to the general dataset that was made available to the public, but cannot be confirmed without access to those data. However, despite these challenges the overall reproducibility of the figure was improved by the suggestive names assigned to each data unique subset when read into the original code. This definitely helped point my efforts in the right direction to recreate the unique sets necessary for statistical analysis. I believe that if the right data is made available, this paper would be easily reproducible for anyone.

Task 2)

2. My second task was an attempt to reproduce Figures 4 from the same paper, this time using laboratory experimental data (lab_expt) collected by the authors. Figure 4 evaluates differences in average soldier deployment to new entrances under high or low threat conditions, simulated using a natural nest competitor of the turtle ant. Initially, I wanted to try with Figure 5, a linear regression, as well, but faced similar challenges as the first task due to the multiple statistical steps which precede this figure. The external datasets used here seem to be manipulated to a higher extent than previous. At this point in time, gaining practical knowledge of the stat functions used is key to adequately reproducing published figures.

3. The data accessed from the web reads into R with no problem. The other data sets referenced in the code are not available. I attempt to recreate the groupings here.

```r
et_sum <- as.factor(Lab_expt$colony)
treatment <- as.factor(Lab_expt$treatment)
et_sum_sizes <- as.vector(Lab_expt$entrance)

## CAVITY USE & DEFENSE
### Subset data based on what steps are taken by Dr. Powell in the original code.
et_sum_nt<- (Lab_expt[treatment=="no_threat",])
et_sum_t<-(Lab_expt[treatment=="threat",])

et_sum_sizes_l<-(et_sum_sizes["l"])
et_sum_sizes_s<-(et_sum_sizes["s"])


### Figure for level of defense in small and large cavities in two contexts. The chunk below is almost

library(lattice)
dotplot(jitter(soldiers, 0) ~ entrance, group=colony, data= et_sum_nt, col="black", type=c("b"),
        ylab="Mean soldiers per cavity", xlab="Entrance size", aspect = 3/1.1, ylim=c(0.6,8.4))
```
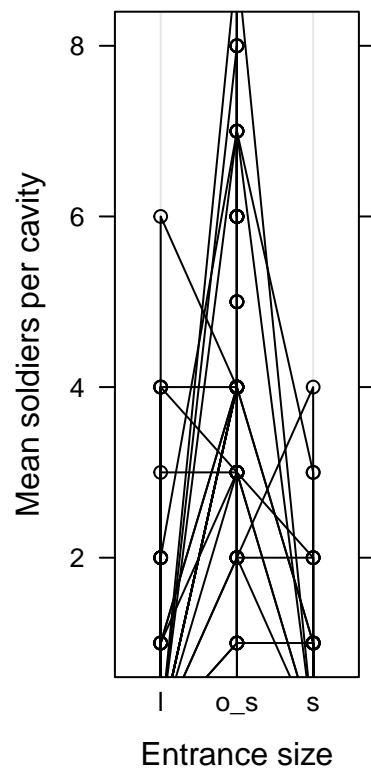
3
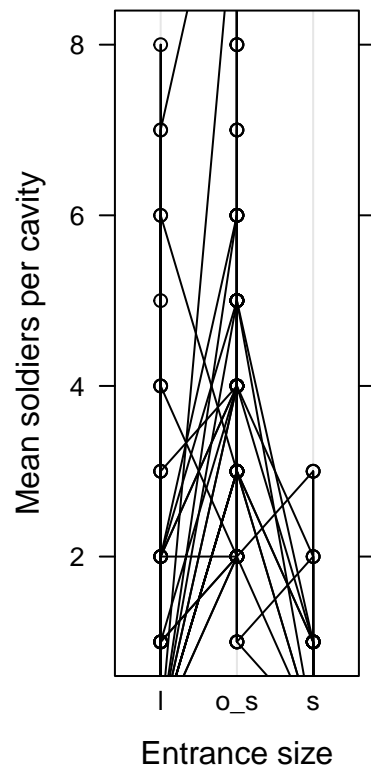
```
library(lattice)
dotplot(jitter(soldiers, 0) ~ entrance, group=colony, data= et_sum_t, col="black", type=c("b"),
        ylab="Mean soldiers per cavity", xlab="Entrance size", aspect = 3/1.1, ylim=c(0.6,8.4))
```

5. Though not successfully completed, this task taught be alot about how variable the structure of code can be. I was able to familiarize myself with a new package, 'lattice', which is primarily used to improve the base R graphics for demonstrating multivariate relationships. Trying to use this new package proved more difficalt than expected, and so I was unable to pinpoint how exactly to remove the middle entrance size category (on x axis, entrance size) or whether that was another consequence of the external datasets used by the author. Because this is a computer language, there are many ways of going about the same task- so even though I could not reach the final figure, the comments provided on the code helped outline the general flow of steps required. Not having the exact data that is read into the original code was effectively the only barrier to reproducing the figure, but I can see from the original data file names that a deep-seated grasp of the study system and methodology employed (which is innate for the original researchers) makes it easy to recognize which variables needed to be summarized for a particular function. Even without providing the extraneous datasets or metadata, the reproducibility of this paper would have been improved by a comment section explaining how the general data needed to be manipulated at each phase or context as to why he brings in so many distinct sets. The results from analyses including T tests and regressions are included in the original code, which is definitely a helpful aspect to include. It is even written in the paper, "The laboratory experiments were designed to allow simple but powerful paired comparisons between metrics of soldier deployment following the treatments" (Powell et AL. 2017, pp. 2291). However, it is important to note that this code was written in 2016-17. An interesting project would be to attempt reproducing one of Dr. Powell's more recent papers from 2019-2020, and compare the degree of reproducibility between the two. Similarly, this task may have been enhanced if I had shifted entirely to recreating the extra data sets themselves, which could simply be checked by running through the original code.

6. This project has taught me alot about the scientific process as a whole, especially reminding me that there is a clear gap between understanding the complex mechanisms behind your work and having the capability to explain the processes to someone not involved or lacking the same experience as you. Modern scientific development appears not in the technological advancements (which I would argue is a constant feature of science today) but as we supersede historical barriers to entry. The advancement of scientific research thus comes in two parts, first the exploration and conclusions drawn from ones own research, and second being the implementation and dissemination of the gained knowledge.

Powell, Scott, Matina Donaldson-Matasci, Augustus Woodrow-Tomizuka, and Anna Dornhaus. 2017. "Context-Dependent Defences in Turtle Ants: Resource Defensibility and Threat Level Induce Dynamic Shifts in Soldier Deployment." Journal Article. *Functional Ecology* 31 (12): 2287–98. https://doi.org/10.1111/1365-2435.12926.