frontiers

Understanding differences between continuous and discrete distributions for statistical models in ENVI200+ ANU course

Samual Smeal [1], Anthony Davidson [2], Course Lecture (coming) [1,3*]

[1] *School of Biology, ANU, Institution X, City X, State XX, Country X*
[2] *Department X, Institution X, City X, State XX, Country X*

Correspondence*:
Course Lecture (coming)
anthony.davidson@anu.edu.au

ABSTRACT

Abstract length and content varies depending on article type. Refer to http://www.frontiersin.org/about/AuthorGuidelines for abstract requirement and length according to article type.

Understanding the population dynamics of a population can be vital for understanding the projection of these populations to different states of abundance and growth.

"*Polytelis swainsonii* – Superb Parrot are a non-excavating species and rely on the presence of tree hollows in a given area to make nests. The species are in what is known to be an extinction debt and without the recruitment of new trees amongst the landscape, the superb parrot will be in an expected decline. To increase the chances of incorporating land management practices, key characteristics based on nest selection of the Superb Parrot will be identified using two main methods including single rope climbing techniques and ground based observational methods. I then used this data to find whether tree physiognomy can be used to predict the number of active nests and the presence of hollows. The Study site is based in surrounding Canberra, a key site to mimic a landscape with little to no recruitment of new trees. Highlighting current and future obstacles regarding the management of hollow bearing woodlands seen across most superb parrot nesting populations I found that climbing data is more accurate compared to ground-based methods in counting tree hollows and tree hollow nests. On average climbing data increased the ground count by 32(insert data here maybe) Super parrots prefer a tree DBH on average 114cm to hollows without a nest 102cm.

Using a double sampling technique paired with a critical ratio, climbing and ground based results in ongoing studies can further contribute to this area of study by replicating the process used in this report. Citizen science projects can benefit in the future by focusing on the key indicators highlighted and thus increase the effectiveness across future measurements. citizen science projects can measurements can be calibrated using double sampling techniques to correct data derived from future citizen science projects, savings costs overall and improving management studies by increasing the observability of simple measurements such as tree DBH"

Keywords: Biology, Statistics, Ecology, Simpsons paradox

INTRODUCTION

Individual biases in the scientific endenvour is a fundamental task of any researcher. This publication links the open-sources tools and templates that allow to for the production of typed set articles.

Here is an example of a simple journal submission template for the journal "Frontiers in Science" and the association information needed (LastName1 et al., 2013). As demonstrated in OtherAuthor and Coauthor (2012), citations can also be automatically reference. Multiple references are separated by semicolons (LastName1 et al., 2013; Author4 and Author5, 2013a). I have used this template to generate a simple description and information about a important issue when comparing means of two samples.

Simpsons paradox (Author4 and Author5, 2013b) is a phenomenon where the true relationship between between two groups is missed due to a underlying, interactions between an unmeasured variable and the two groups being investigated.

(Simpson, 1951)

This can occur and not be know but often it happens that later research can identify these paradoxical relationships with additional data. The study set up for this experiment has the potential for this to occur, particularly because of the modification and manipulation of the original data.

METHODS

"Methods Study area and species The data collected was from North Canberra in two sites, the outskirts of Throsby just outside of mulligans flat reserve -35.172074, 149.165799 and south west of Belconnen -35.241217, 149.004370 where the superb parrot is known to breed in colonies on the outskirts of Canberra. The two study sites are roughly 15km apart on either side of suburban Canberra and can be considered independent of each other with localised climate history and also dependent on preference due to the ability of the superb parrots metapopulation not being limited to the distance between the study sites. The study sites are heavily grazed open woodlands used for mostly cattle. The trees are mostly all aged hollow bearing little to no recruitment of young trees sue to the current land uses. For this study sites measurements were derived by only two tree species, Eucalytpus blakleyi (Blakely's red gum) and Eucalyptus rosii (Scribbly Gum)"

0.0.1 Continuous variables

A continuous variable is estimated with a mean and SD and contains values that include all real numbers. The first data set has the following arrangement of continuous and discrete variables.

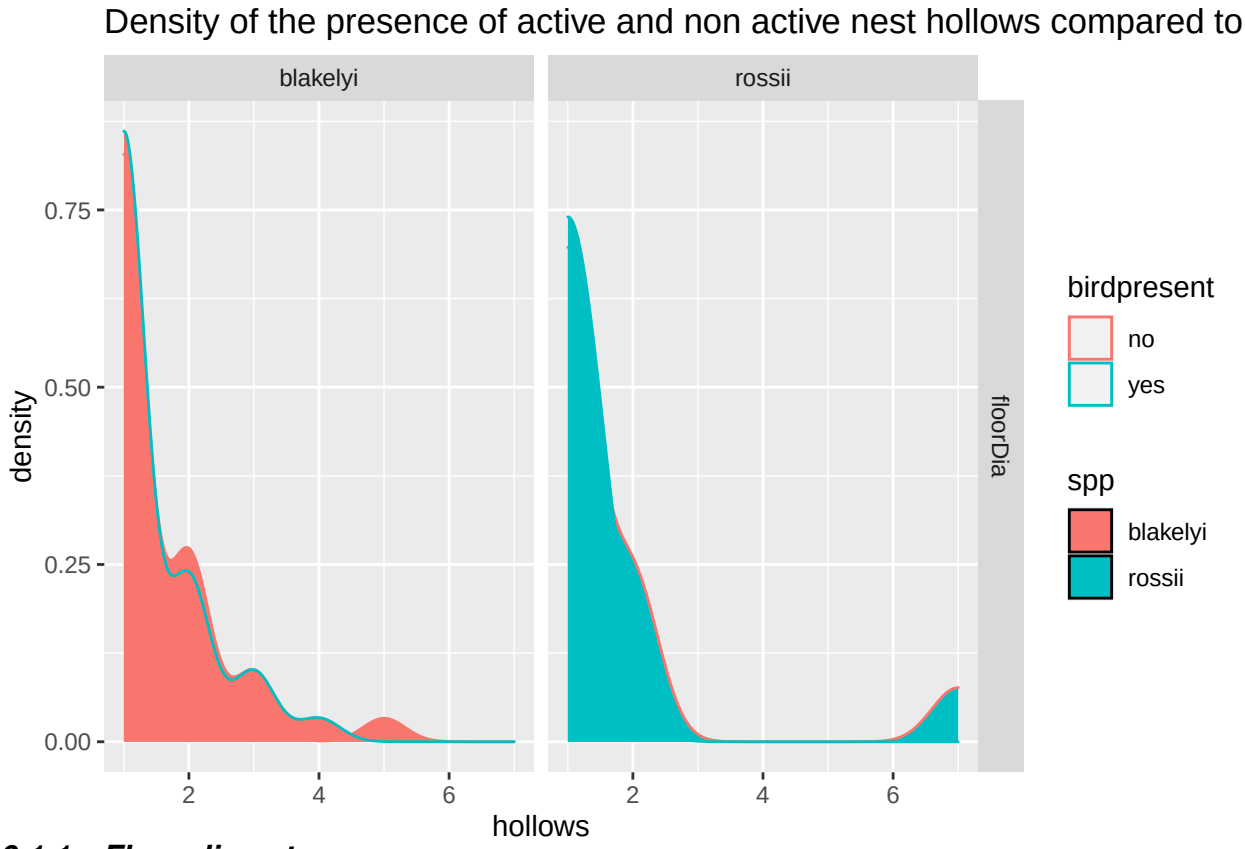
```
#data
dat_hol <- read_xlsx(here::here("SamsRepo/data/hollow_data_wrangled.xlsx"),sh
mutate(BirdPresence = factor(birdPresent),
spp = factor('tree species'))

kableExtra::kable(glimpse(head(dat_hol)))
```

```
## Rows: 6
## Columns: 9
## $ birdPresent                                <chr> "no", "no", "no", "no
## $ `number of entrance holes per hollow`      <dbl> 1, 1, 2, 2, 1, 1
## $ `minimum entrance diameter (cm)`           <dbl> 6, 9, 10, 13, 4, 4
```

```
## $ `floor diameter (cm)`                                <dbl> 8, 9, 10, 26, 5, 8
## $ `depth of hollow (cm)`                               <dbl> 61, 19, 72, 70, 35, 5
## $ `diameter of the stem the hollow is in (cm)` <dbl> 17, 15, 18, 43, 91, 1
## $ `tree species`                                       <chr> "blakelyi", "blakelyi
## $ BirdPresence                                         <fct> no, no, no, no, no, n
## $ spp                                                  <fct> tree species, tree sp
```

birdPresent	number of entrance holes per hollow	minimum entrance diameter (cm)	floor diameter (cm)	c
no	1	6	8	
no	1	9	9	
no	2	10	10	
no	2	13	26	
no	1	4	5	
no	1	4	8	

```
varnames_hol <- read_xlsx(here::here("SamsRepo/data/hollow_data_wrangled.xlsx

names(dat_hol)[1:7] <- varnames_hol$shortname

dat_hol <- dat_hol[,1:7]

dat_hol %>%
  ggplot(mapping = aes(x = hollows, col = birdpresent, fill = spp)) +
  # geom_histogram(position = "dodge", stat = "count") +
  geom_density() +
  facet_grid("floorDia"~spp) +
  coord_cartesian() +
  ggtitle("Density of the presence of active and non active nest hollows comp
```



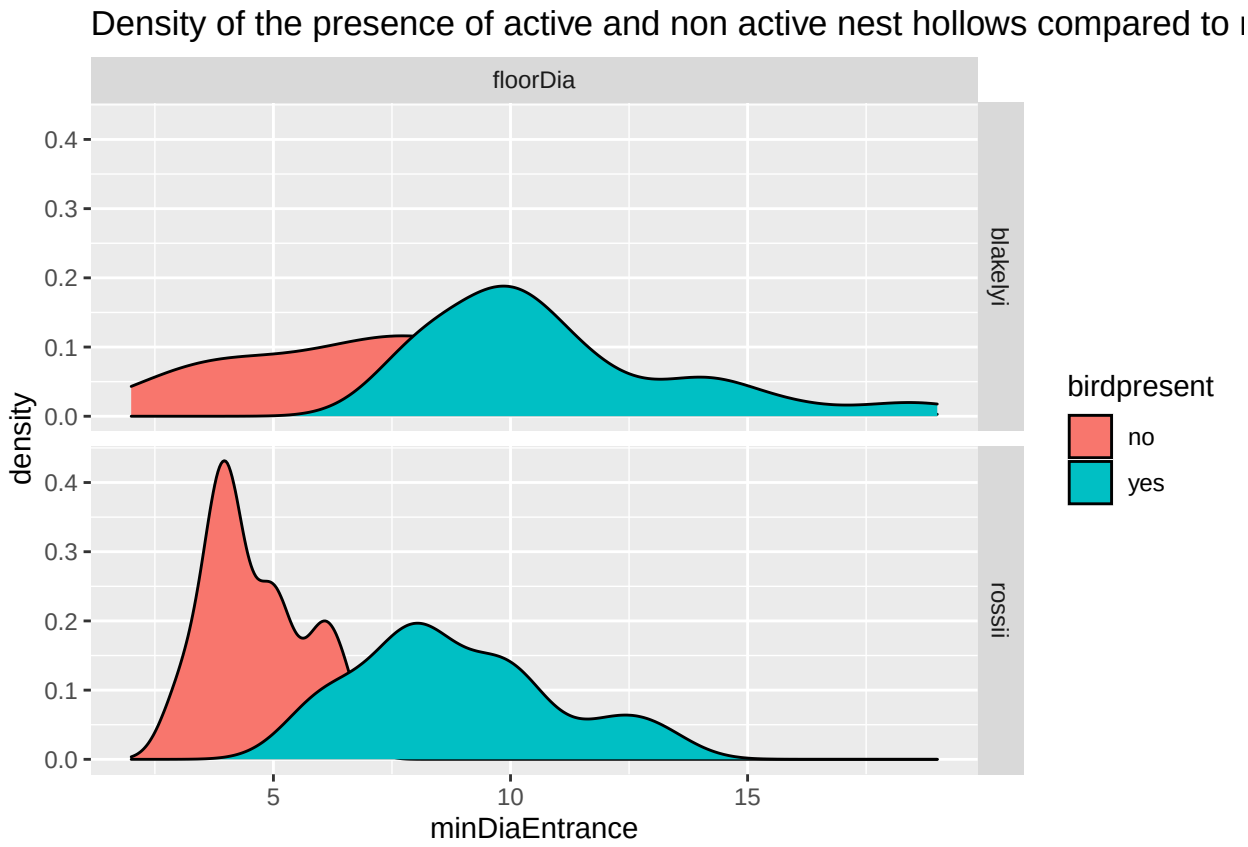
0.0.1.1 Floor diameter

```
names(dat_hol)
```

```
## [1] "birdpresent"    "hollows"        "minDiaEntrance" "floorDia"
## [5] "depthCm"        "diaCm"          "spp"
```

```
dat_hol %>%
  ggplot(mapping = aes(x = minDiaEntrance, fill = birdpresent)) +
  # geom_histogram(position = "dodge", stat = "count") +
  geom_density(bins = 100) +
  facet_grid(spp~"floorDia") + #
  coord_cartesian() +
  ggtitle("Density of the presence of active and non active nest hollows comp
```

```
## Warning: Ignoring unknown parameters: bins
```



0.0.1.2 Normal

```
names(dat)
```

```
## [1] "BirdPresence"    "treeDiameter"    "healthScale"     "hollowNumber"
## [5] "hollowEntrances" "spp"
```

```
glimpse(dat)
```

```
## Rows: 50
## Columns: 6
## $ BirdPresence    <fct> no, no, no, no, no, no, no, no, no, no, no, no, no
## $ treeDiameter    <dbl> 125, 100, 106, 136, 70, 113, 95, 83, 67, 48, 96, 9
## $ healthScale     <dbl> 4, 2, 3, 4, 4, 4, 2, 3, 2, 4, 2, 4, 1, 1, 2, 3, 3,
## $ hollowNumber    <dbl> 5, 1, 11, 9, 2, 13, 4, 4, 2, 0, 3, 1, 3, 18, 6, 5,
## $ hollowEntrances <dbl> 3, 2, 3, 7, 2, 14, 3, 2, 1, 1, 0, 3, 3, 9, 7, 3, 2
## $ spp             <fct> blakelyi, blakelyi, blakelyi, blakelyi, blakelyi,
```

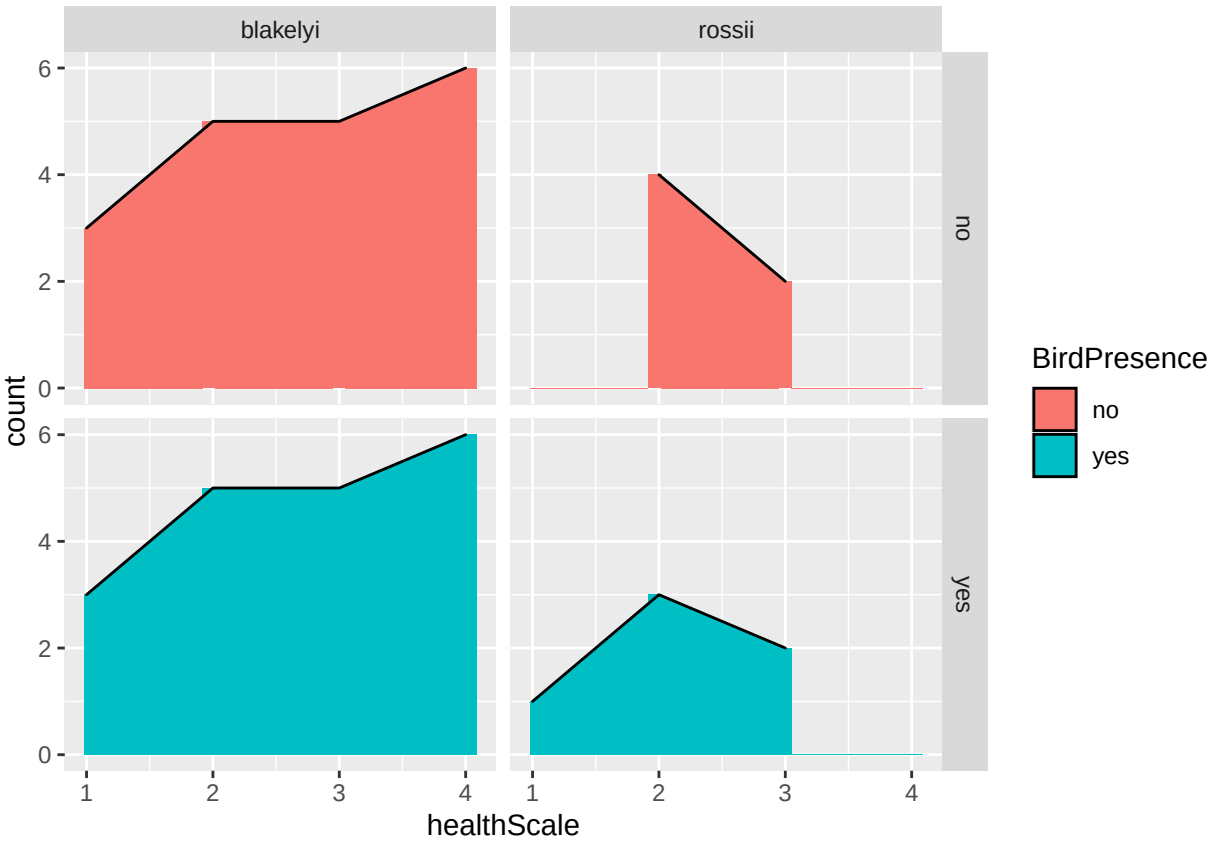
```
dat1 <- dat[,2,5]
# plot(lm(select(dat, treeDiameter, healthScale), subset = dat$BirdPresence))

varNames <- names(dat)
```

```
varLength <- length(names(dat))
# for(i in 1:varLength) {
#   i = 2
#   mod <- lm(select(dat, varNames[i]))
#
#   ggplot(dat) +
#     geom_histogram(aes(x = dat$varNames[i]), stat = "count")
# }
```

- Tree Diameter

```
dat %>%
  ggplot(mapping = aes(healthScale, fill = BirdPresence)) +
  geom_histogram(position = "dodge") +
  geom_density(stat = "count") +
  facet_grid(BirdPresence~spp) +
  coord_cartesian()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

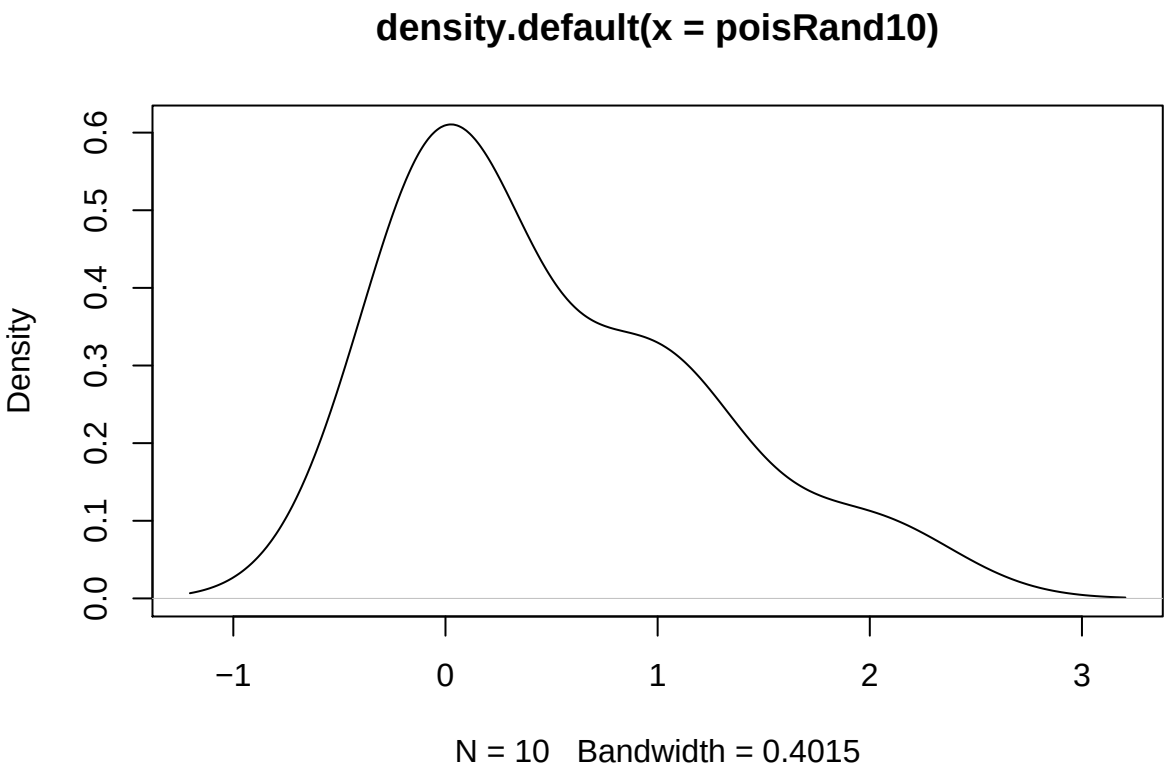


```
#random health scale?
names(dat)

rand_scale <- filter(dat, BirdPresence == "no")


tallyCounts <- dat %>%
                group_by(BirdPresence) %>%
                  summarise(treeDiameter = mean(treeDiameter),
                            healthScale = mean(healthScale),
                            hollowEntrances = mean(hollowEntrances),
                            hollowNumber = mean(hollowNumber),
                            Random = mean(as.numeric(BirdPresence)),
                            spp = mean(as.numeric(spp)))

tallyCounts

y <- dat$BirdPresence
x <- dat$healthScale

datNull <- data.frame(y,x)
#anova in R
#is just glm then avo of output

mod <- glm(y ~x, data = datNull, family = "binomial")
summary(mod)
anovaR <- anova(mod)

health_scale
```

0.0.1.3 Poisson

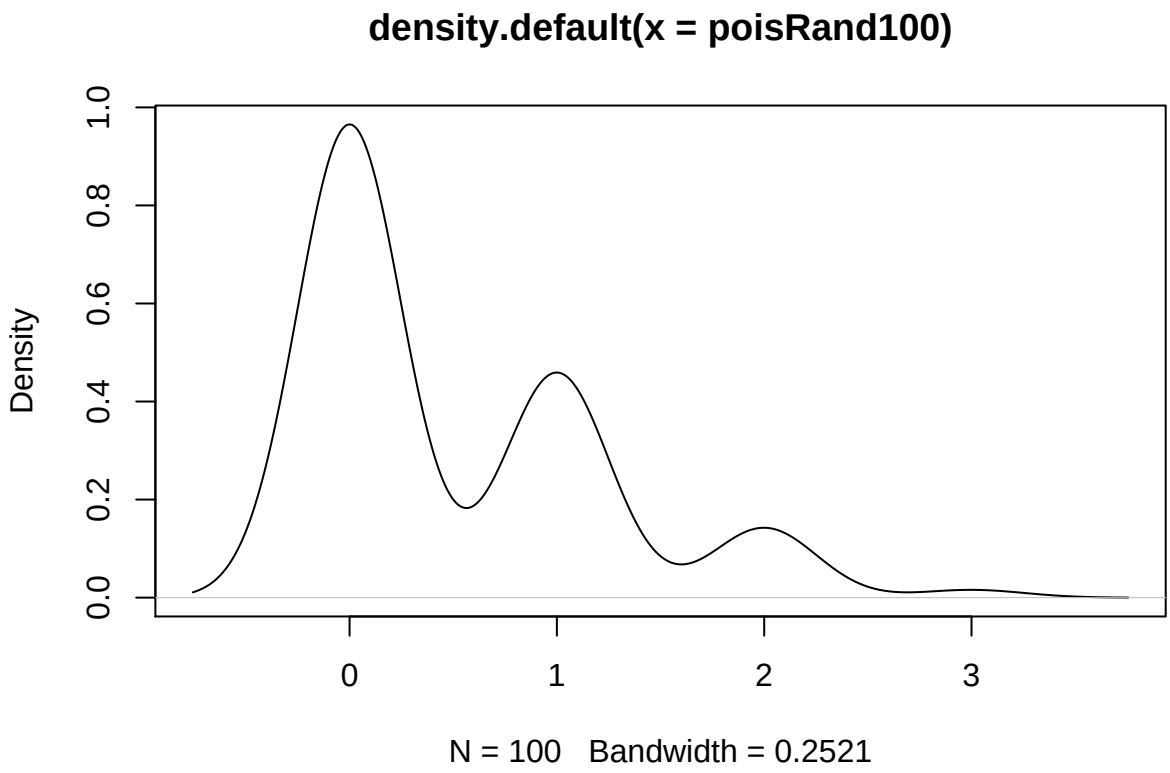
- hollowNumber
- hollowEntrances

```
poisRand10 <- rpois(10, 0.5) #lambda is growth rate here
plot(density(poisRand10))

poisRand20 <- rpois(20, 0.5) #lambda is growth rate here
plot(density(poisRand10))
```



92

```
poisRand30 <- rpois(30, 0.5) #lambda is growth rate here
plot(density(poisRand10))

poisRand100 <- rpois(100, 0.5) #lambda is growth rate here
plot(density(poisRand100))
```



0.0.1.4 Discrete

- BirdPresence
- spp

```
# DT::datatable(dat)
kableExtra::kable(head(dat))
```

BirdPresence	treeDiameter	healthScale	hollowNumber	hollowEntrances	spp
no	125	4	5	3	blakelyi
no	100	2	1	2	blakelyi
no	106	3	11	3	blakelyi
no	136	4	9	7	blakelyi
no	70	4	2	2	blakelyi
no	113	4	13	14	blakelyi

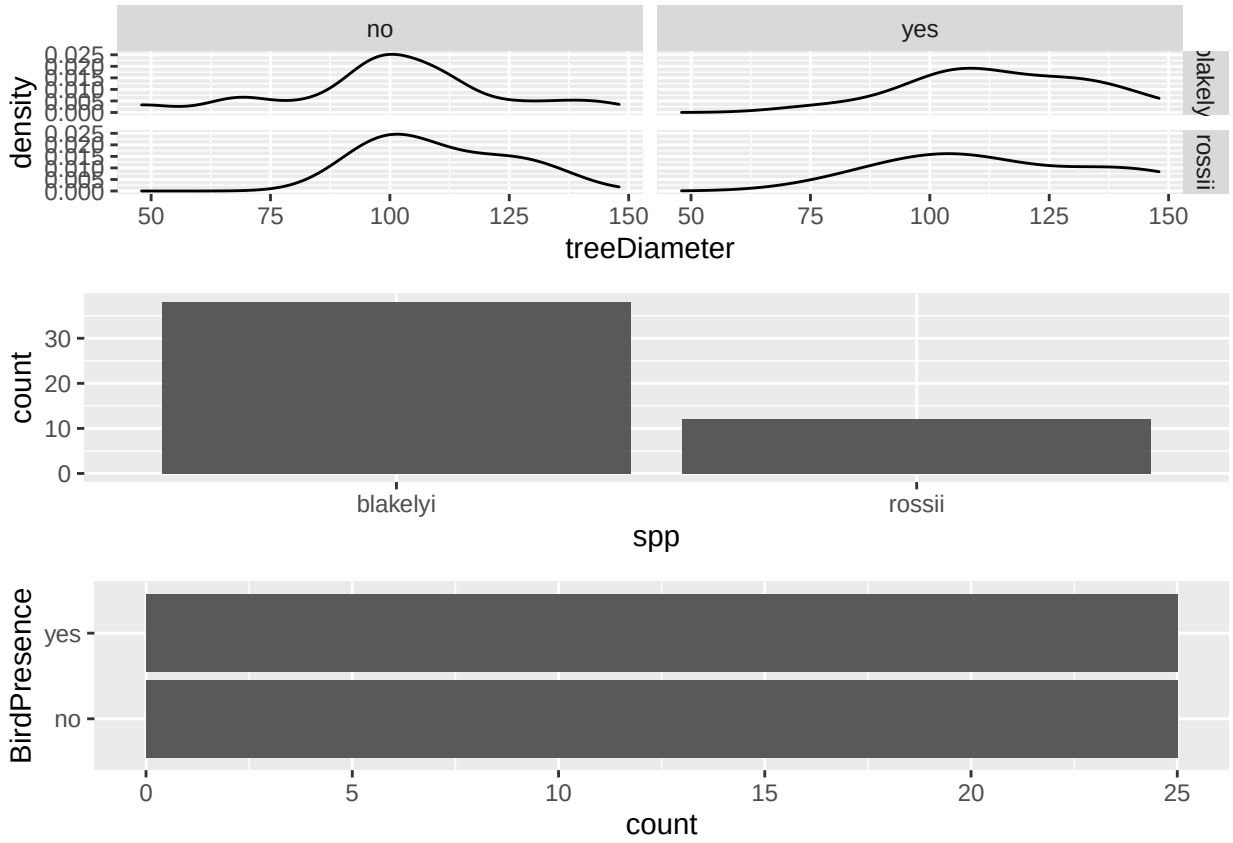
RESULTS

```
p1 <- dat %>%
  ggplot(mapping = aes(treeDiameter, group = BirdPresence)) +
  geom_density() +
  facet_grid(spp~BirdPresence)

p2 <- dat %>%
```

```
  ggplot(mapping = aes(spp)) +
  geom_histogram(stat = "count") +
  coord_cartesian()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

```
p3 <- dat %>%
  ggplot(mapping = aes(BirdPresence)) +
  geom_bar() +
  coord_flip()

gridExtra::grid.arrange(p1,p2,p3)
```




```
# barplot_test <- plot(x = dat$healthScale, y = dat$BirdPresence)
# str(dat$BirdPresence)
```

Hollow data

coming???

```
#balanced hollow
dat_holl <- read_xlsx(here::here("SamsRepo/data/hollow_data_wrangled.xlsx"),s
varnames_holl <- read_xlsx(here::here("SamsRepo/data/hollow_data_wrangled.xls
```

Frontiers requires figures to be submitted individually, in the same order as they are referred to in the manuscript. Figures will then be automatically embedded at the bottom of the submitted manuscript. Kindly ensure that each table and figure is mentioned in the text and in numerical order. Permission must be obtained for use of copyrighted material from other sources (including the web). Please note that it is compulsory to follow figure instructions. Figures which are not according to the guidelines will cause substantial delay during the production process.

1 DISCUSSION

The results of this publication highlight the ability for any undergraduate course to include the relevent documentation for the submission of research articles from written text from a word document from an RMarkdown veriation of this information.

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

The statement about the authors and contributors can be up to several sentences long, describing the tasks of individual authors referred to by their initials and should be included at the end of the manuscript before the References section.

ACKNOWLEDGMENTS

Funding:

2 SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder

3 REFERENCES

A reference list should be automatically created here. However it won't. Pandoc will place the list of references at the end of the document instead. There are no convenient solution for now to force Pandoc to do otherwise. The easiest way to get around this problem is to edit the LaTeX file created by Pandoc before compiling it again using the traditional LaTeX commands.

FIGURES

Author4, N., and Author5, N. (2013a). Title of the article. *Frontiers in Neurorobotics* 7. doi:10.3389/fnbot.2013.56789.

Author4, N., and Author5, N. (2013b). Title of the article. *Frontiers in Neurorobotics* 7. doi:10.3389/fnbot.2013.56789.

LastName1, A., LastName2, A., and LastName2, A. (2013). Article title. *Frontiers in Neuroscience* 30, 10127–10134. doi:10.3389/fnins.2013.12345.

OtherAuthor, N., and Coauthor, N. S. (2012). Article title. *Frontiers in Genetics* 30, 16417–16418. doi:10.3389/fgene.2012.54321.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 13, 238–241.