

Assignment 7

2022-03-25

```
library(readr)
library(ggplot2)
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
```

Read the daphnia data set

```
daphnia <- read_csv("Daphniagrowth.csv")

## Rows: 40 Columns: 3
## — Column specification
## Delimiter: ","
## chr (1): parasite
## dbl (2): rep, growth.rate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

#Test differences among multiple groups (only #two groups are compared in a t-test)

```
daphnia <- read_csv("daphniagrowth.csv")

## Rows: 40 Columns: 3
## — Column specification
## Delimiter: ","
## chr (1): parasite
## dbl (2): rep, growth.rate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
cols(
  parasite = col_character() ,
  rep = col_double() ,
  growth.rate = col_double ())

## cols(
##   parasite = col_character(),
##   rep = col_double(),
##   growth.rate = col_double()
## )

tbl_df(daphnia)

## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## # A tibble: 40 × 3
##   parasite    rep growth.rate
##   <chr>    <dbl>      <dbl>
## 1 control      1      1.07
## 2 control      2      1.27
## 3 control      3      1.32
## 4 control      4      1.08
## 5 control      5      1.20
## 6 control      6      1.38
## 7 control      7      1.32
## 8 control      8      1.10
## 9 control      9      1.28
## 10 control     10      1.12
## # ... with 30 more rows
```

#Biological Question: Do parasites impede growth of water flea? #Statistical Null Hypothesis (H0): No effect (equal μ , the mean #growth rates among groups)

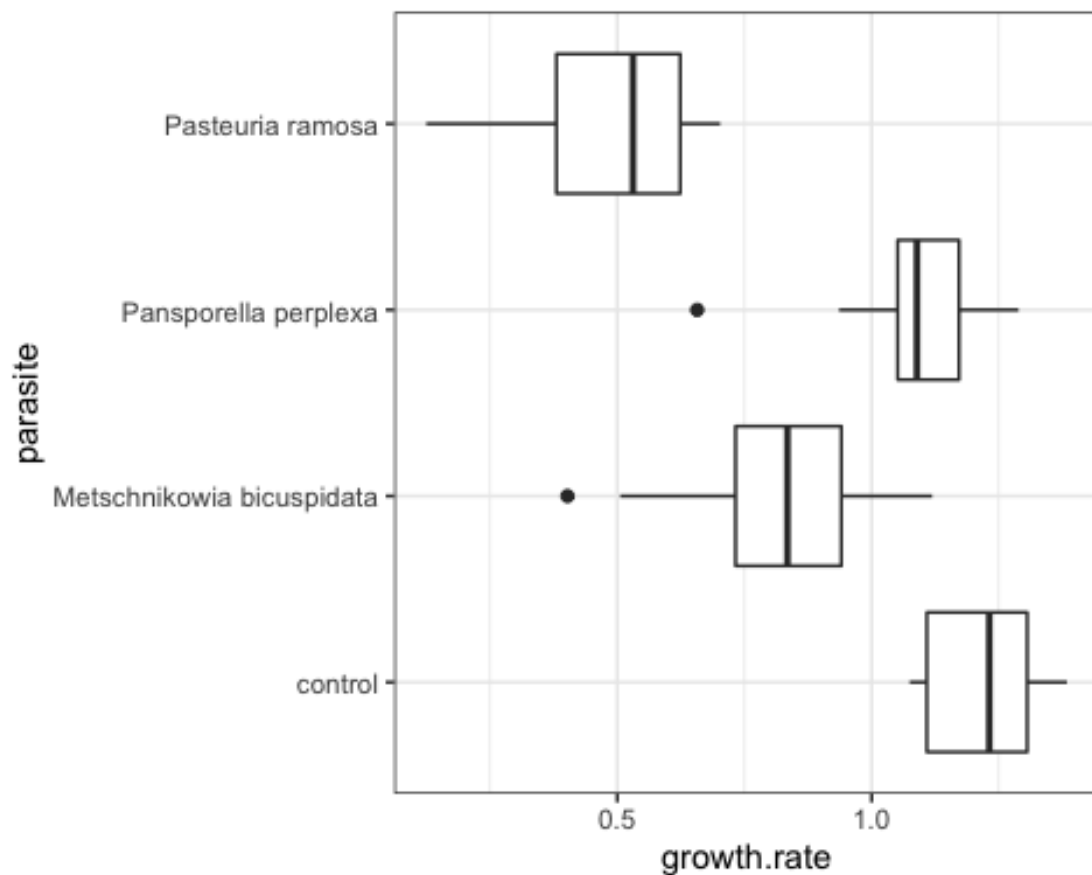
```
glimpse(daphnia)

## Rows: 40
## Columns: 3
## $ parasite    <chr> "control", "control", "control", "control", "control",
## "co...
## $ rep         <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1, 2, 3, 4, 5, 6, 7, 8,
## 9, ...
## $ growth.rate <dbl> 1.0747092, 1.2659016, 1.3151563, 1.0757519, 1.1967619,
## 1.3...
```

#One-Way ANOVA. Step II. Visualize with a box plot

```
ggplot(daphnia, aes(x = parasite,
  y = growth.rate)) +
```

```
geom_boxplot() +  
theme_bw() +  
coord_flip()
```



#Calculate mean growth rates by #groups (the “parasite” column). Save #to a new object “sumDat”, with the #mean growth rates in a column #named as “meanGR”.

```
sumDat <- daphnia %>% group_by(parasite) %>% summarise(meanGR =  
mean(growth.rate))  
sumDat
```

```
## # A tibble: 4 × 2  
##   parasite          meanGR  
##   <chr>          <dbl>  
## 1 control          1.21  
## 2 Metschnikowia bicuspidata 0.801  
## 3 Pansporella perplexa      1.08  
## 4 Pasteuria ramosa         0.482
```

```
model_grow <- lm(data = daphnia, growth.rate ~ parasite)
```

#One-Way ANOVA. Step III. Test with a linear model

```

model_grow <- lm(data = daphnia, growth.rate ~ parasite)
anova(model_grow)

## Analysis of Variance Table
##
## Response: growth.rate
##           Df Sum Sq Mean Sq F value    Pr(>F)
## parasite   3  3.1379  1.04597   32.325 2.571e-10 ***
## Residuals  36  1.1649  0.03236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model_grow)

##
## Call:
## lm(formula = growth.rate ~ parasite, data = daphnia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41930 -0.09696  0.01408  0.12267  0.31790
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21391    0.05688   21.340 < 2e-16 ***
## parasiteMetschnikowia bicuspidata -0.41275    0.08045   -5.131 1.01e-05 ***
## parasitePansporella perplexa      -0.13755    0.08045   -1.710  0.0959 .
## parasitePasteuria ramosa          -0.73171    0.08045   -9.096 7.34e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1799 on 36 degrees of freedom
## Multiple R-squared:  0.7293, Adjusted R-squared:  0.7067
## F-statistic: 32.33 on 3 and 36 DF,  p-value: 2.571e-10

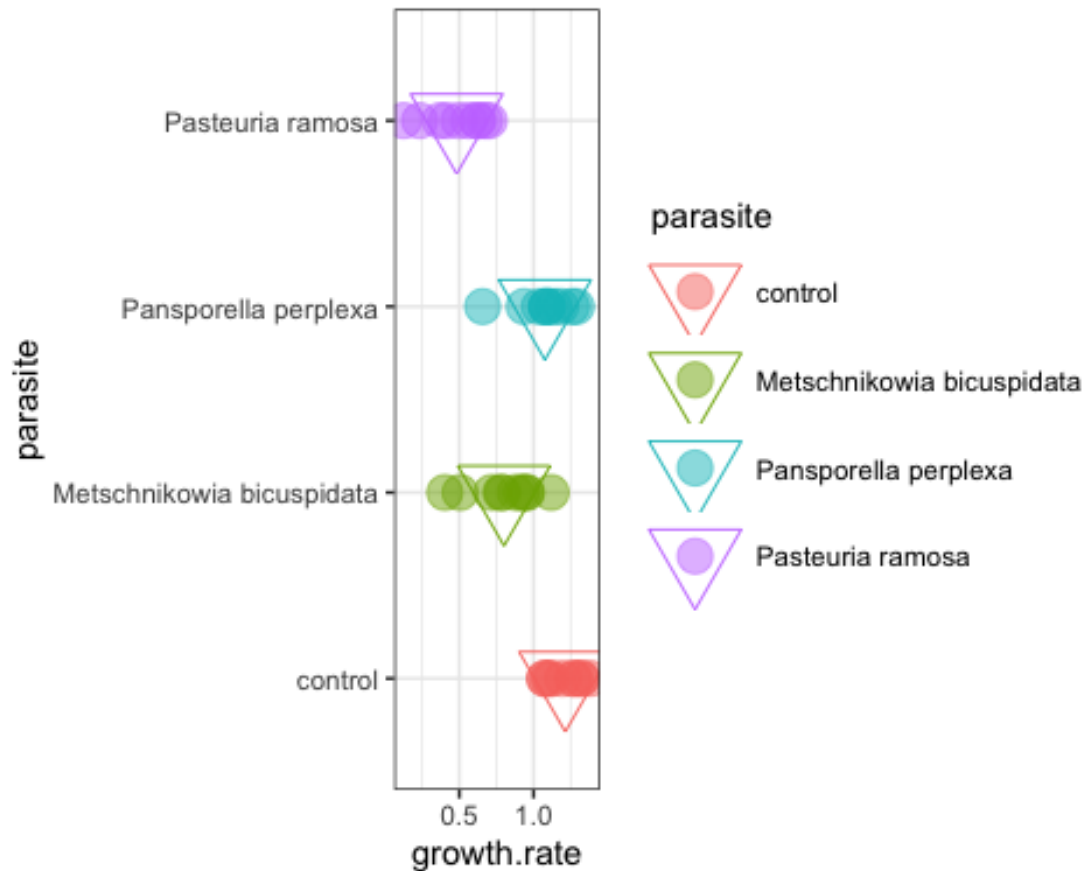
```

#One-Way ANOVA IV. Estimate parasite effects & replot

```

ggplot(data = daphnia, aes(x = parasite, y = growth.rate, color =
parasite) ) + geom_point(size = 5, alpha = 0.5) + geom_point(data =
sumDat, aes(x = parasite, y = meanGR, color = parasite), shape=6,
size=10) + theme_bw() + coord_flip()

```



χ^2 Contingency Test I. Load data & Define hypothesis #libraries I always use

```
library(dplyr)
library(ggplot2)
```

#import the data

```
lady <- read.csv("ladybirds_morph_colour.csv")
```

#check it out #Test association of two variables, both of which are #counts (compare with regression, which test #correlation between two continuous variables) #Biological Question: Are habitat and body color of ladybirds associated? #Statistical Null Hypothesis (H0): No association

```
glimpse(lady)
```

```
## Rows: 20
## Columns: 4
## $ Habitat      <chr> "Rural", "Rural", "Rural", "Rural", "Rural", "Rural",
##              "Ru...
## $ Site         <chr> "R1", "R2", "R3", "R4", "R5", "R1", "R2", "R3", "R4",
##              "R5...
## $ morph_colour <chr> "black", "black", "black", "black", "black", "red",
##              "red"...
```

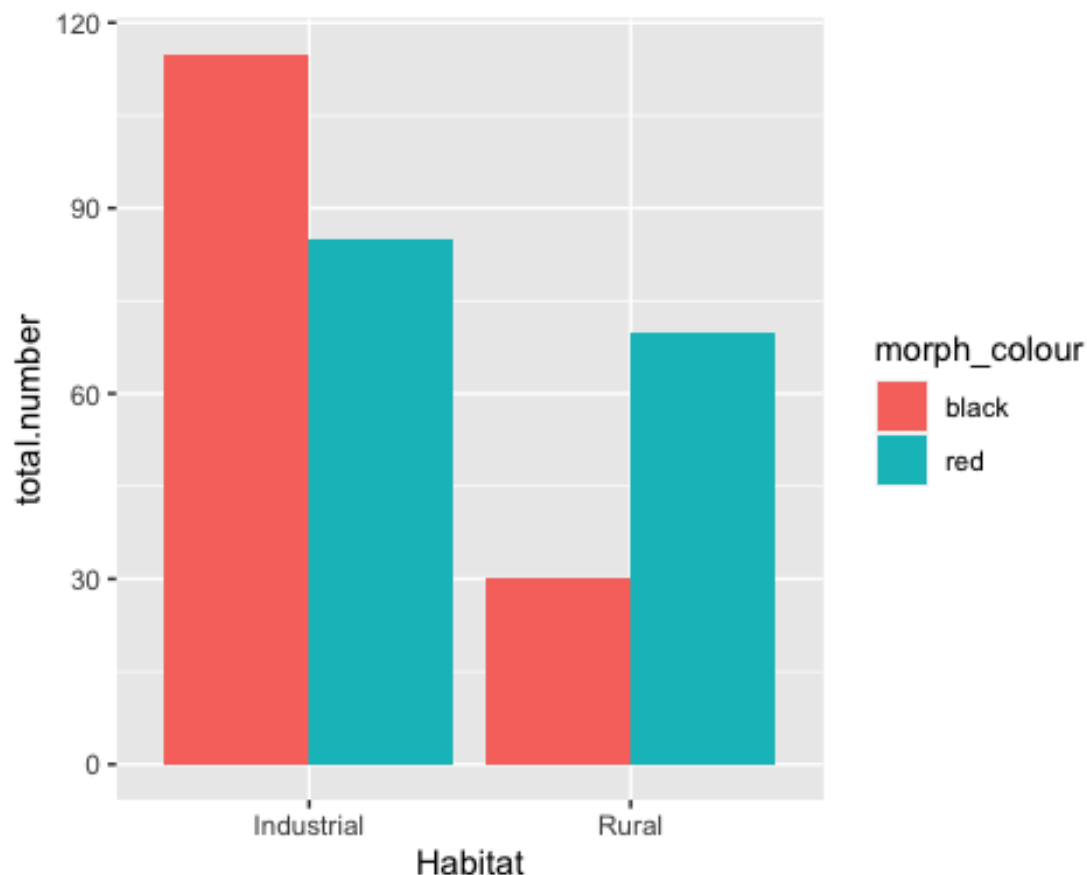
```
## $ number      <int> 10, 3, 4, 7, 6, 15, 18, 9, 12, 16, 32, 25, 25, 17, 16, 17...
```

χ^2 Contingency Test. Step II. Visualize with a bar plot #get counts

```
totals <- lady %>%  
  group_by(Habitat, morph_colour) %>%  
  summarise(total.number = sum(number))  
  
## `summarise()` has grouped output by 'Habitat'. You can override using the  
## `.groups` argument.
```

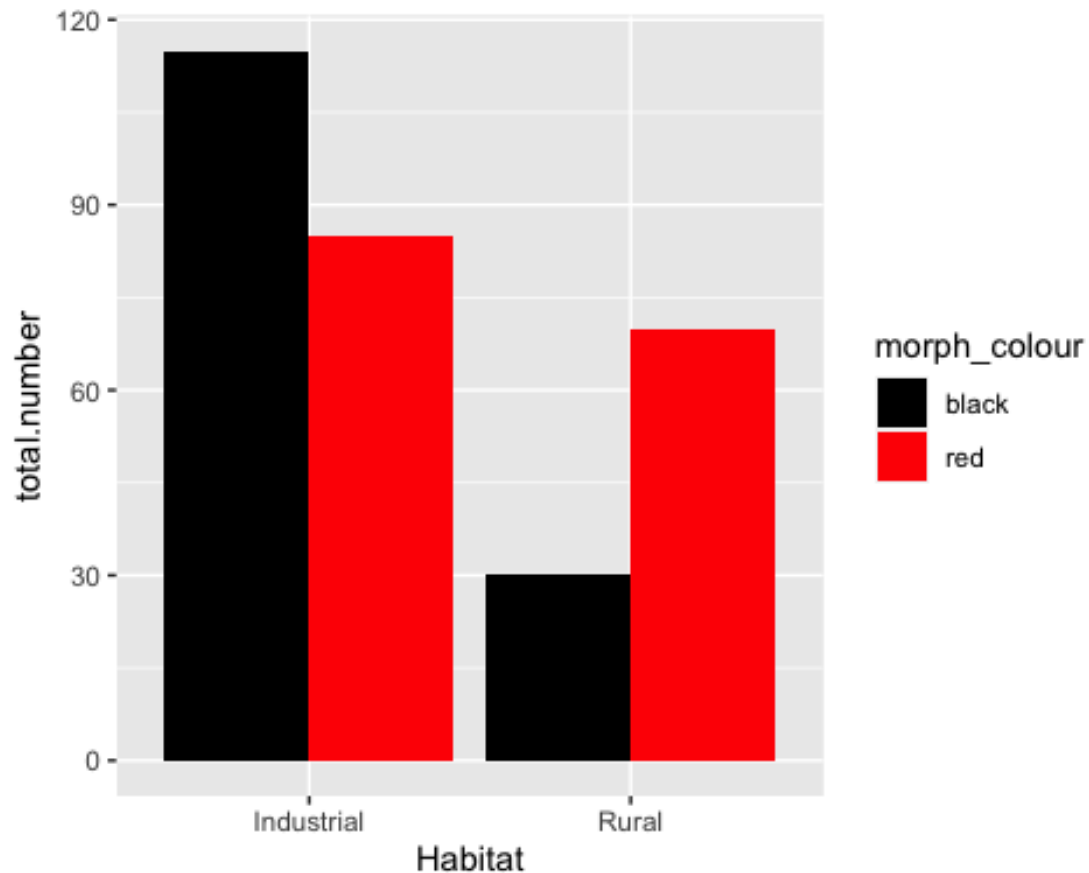
#plot bar graph

```
ggplot(totals, aes(x = Habitat, y = total.number,  
  fill = morph_colour)) +  
  geom_bar(stat = 'identity', position = 'dodge')
```



#customize colors

```
ggplot(totals, aes(x = Habitat, y = total.number, fill = morph_colour)) +  
  geom_bar(stat = 'identity', position = 'dodge') +  
  scale_fill_manual(values = c(black = "black", red = "red"))
```



χ^2 Contingency Test. Step III. Run statistic test & draw conclusions #Black ladybugs are over-represented in industrial #habitats (compare observed vs expected counts)

```
lady.mat<-xtabs(number~Habitat + morph_colour,data=lady)
lady.mat
```

```
##           morph_colour
## Habitat    black red
## Industrial   115  85
## Rural        30  70
```

#obtain p value #'We find insufficient evidence to reject the null hypothesis that Ladybird colour morphs are equally distributed in the two habitats (Chi-sq =, df = , p =)'

```
chisq.test(lady.mat)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  lady.mat
## X-squared = 19.103, df = 1, p-value = 1.239e-05
```