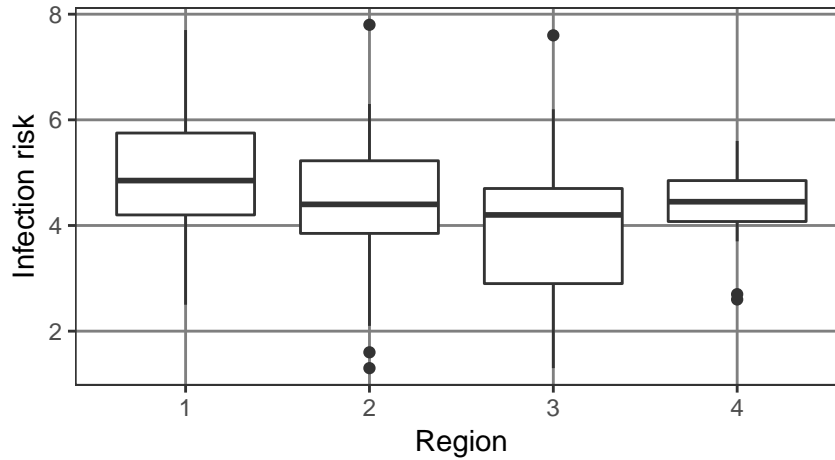
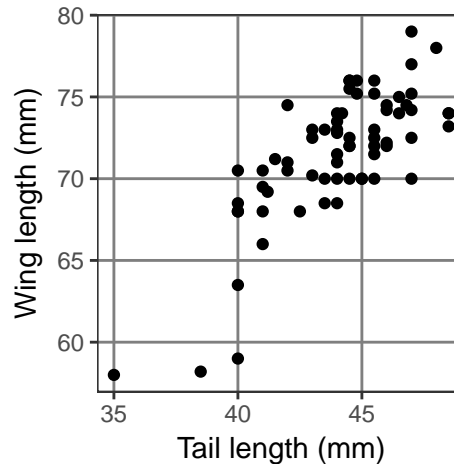


Using the words we learned in this class, answer all questions succinctly; you will lose points for rambling. If you provide R code to answer a problem, provide all the necessary code and indicate clearly which variable(s) contain the answer.

- Below is a boxplot of infection risk by region from a random sample of hospitals from around the United States. Provide the R code to run the **correct** ANOVA of these data.

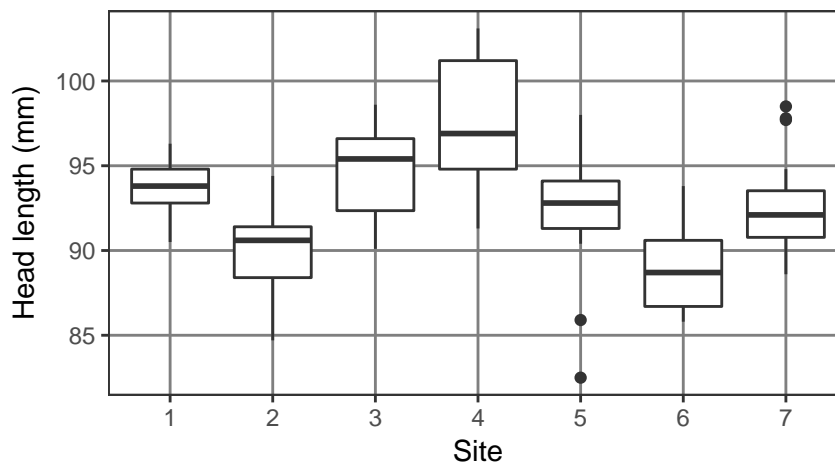


- Consider the variables `taillength` and `winglength`, below, for finches from the Galapagos islands.



- Provide as much R code as you can to mimic the plot above.

- (b) Write R code to make an appropriate 93% confidence interval for these data.
- (c) Set up the appropriately matching hypotheses for the confidence interval above.
- (d) Suppose you computed the interval $(-30.02, -28.04)$. Interpret the provided confidence interval, clearly stating which variable's mean is larger, if any.
- (e) Does the confidence interval provide convincing evidence that there is a real difference in the mean wing to tail length? Explain.
- (f) Explain to my grandmother the statistical conclusion from your confidence interval.
- (g) Would your conclusion differ if instead a 99% confidence interval was used? Explain.
3. Consider a random sample of opossum from multiple sites within Australia. Head length was measured in millimeters.



```
## Analysis of Variance Table
##
## Response: headL
##           Df Sum Sq Mean Sq F value    Pr(>F)
## site         6 499.94  83.323    9.914 1.629e-08 ***
## Residuals  97 815.25   8.405
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

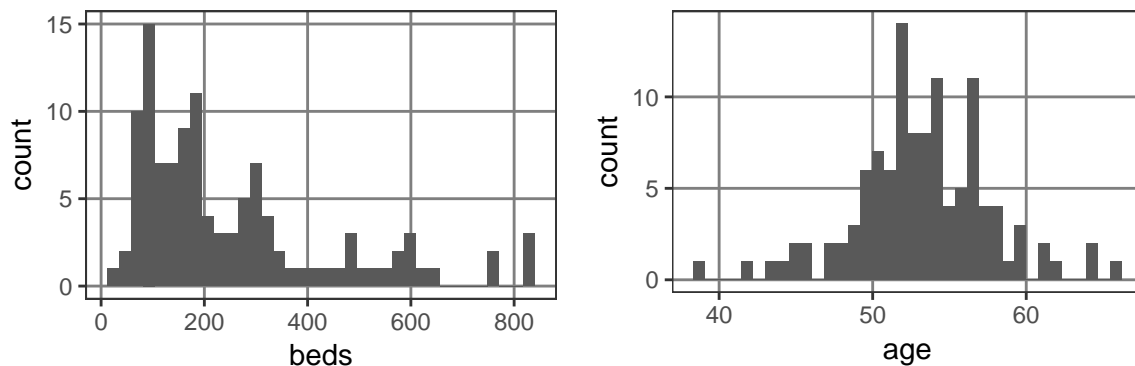
- (a) Provide as much R code as you can to mimic the plot above.
- (b) Set up an appropriate hypothesis test for the analysis above.
- (c) Conclude the hypothesis test.
- (d) Are the assumptions of ANOVA satisfied. Explain each with a complete, English sentence.
- (e) Provide the R code to appropriately test the two groups' means that you believe are most different.
4. Suppose a new graduate is considering a job in two locations, Cleveland, OH and Sacramento, CA, and she wants to see whether the average income in one of these cities is higher than the other. She runs the following R code.

```
t.test(income~city, data=city_incomes, conf.level=0.98)

##
## Welch Two Sample t-test
##
## data:  income by city
## t = -2.4794, df = 99.942, p-value = 0.01483
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 98 percent confidence interval:
## -24133.6883 -573.9385
## sample estimates:
## mean in group Cleveland mean in group Sacramento
##                21627.05                33980.86
```

- (a) Set up, evaluate, and conclude the hypothesis test implicit to the R code above.
- (b) Does the provided confidence interval agree with your conclusion? Explain.
- (c) Explain the literal meaning of a $P\%$ confidence interval.
5. Consider a random of sample of hospitals from the United States, where characteristics of both the hospital and the patients within each hospital were collected. The variable **beds** records the number of beds in the hospital. The variable **age** records average patient age of patients within each hospital.



- (a) Describe the shape of data for each variable above.
- (b) I, Edward, claim that the shape of these data, given their descriptions, was expected. Explain why.