

Introduction to Statistics I

Instructor: Jodin Morey moreyj@lemoyne.edu

Previous Lecture

- ◆ Measures of Spread: Range, IQR, SD
- ◆ Empirical Rule: 1SD: 68%, 2SD: 95%, 3SD: Almost All
- ◆ z-score: $\frac{x - \bar{x}}{s}$



Topic 10: More Summary Measures and Graphs

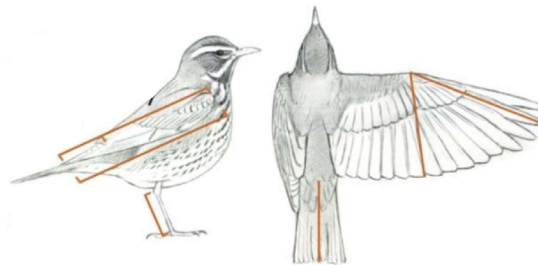


Example: A landmark study on natural selection in 1898 examined sparrows after a particularly severe winter storm. Some sparrows survived and some didn't. Dr. Bumpus measured the size of the sparrows to see if there was a difference for those who survived and those who didn't.



bit.ly/introstatsdata

Polls: Sparrow Size



Data - Sparrows who died:

156	158	160	160	160	161
161	161	161	161	162	162
162	162	162	162	163	163
164	164	165	165	166	166



Max/Min/Median?

Min: 156, Median: 162, Max: 166

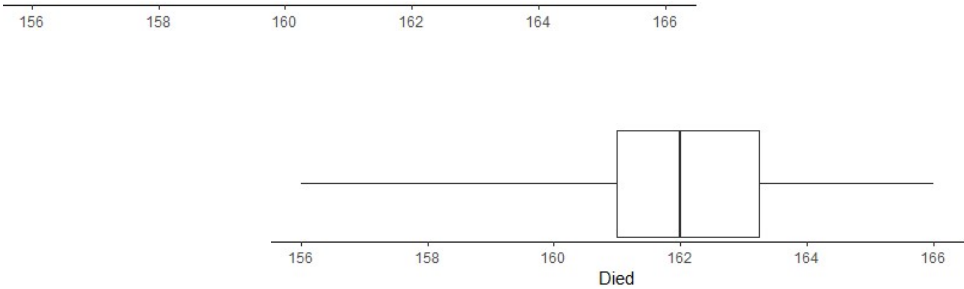
Five Number Summary (FNS) summarizes data: Describes where the four quarters of data fall.

Min, Lower Quartile, Median, Upper Quartile, Max.

Back to Example - FNS:

Min: 156, LQ: 161, Median: 162, UQ: 163.5, Max: 166

How to graph?



Box Plot

Min:156, LQ: 161, Med: 162, UQ: 163.5, Max: 166

Sparrows who Survived

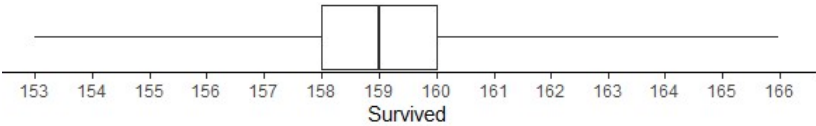
153	154	154	155	156	156	157	157	158
158	158	158	159	159	159	159	159	159
160	160	160	160	160	160	160	160	160
161	161	161	161	162	163	165	166	



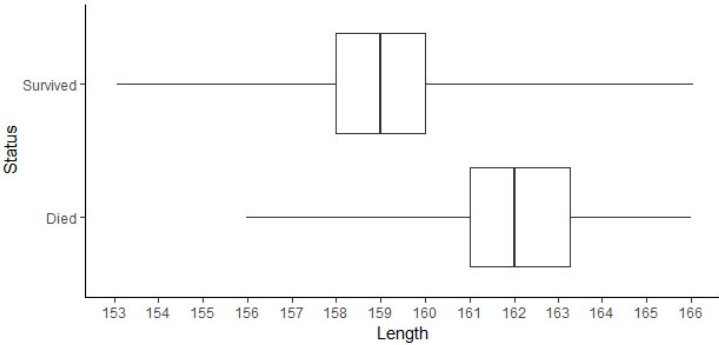
Min/LQ/Median/UQ/Max?

Min: 153, LQ: 158, Median: 159, UQ: 160, Max: 166.

Graph? Compare w/other graph?

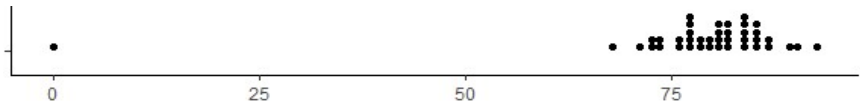


Mn:153, LQ:158, Med:159, UQ:160, Mx:166



Compare

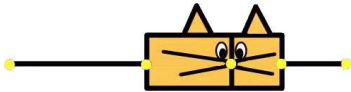
Modified Box Plots & Outliers



Outliers: Max and/or min values may be outliers. We want the distribution excluding outliers.

Need to define what an outlier is, so we can exclude it.

Pts in the box are not outliers. What about pts in the whiskers?

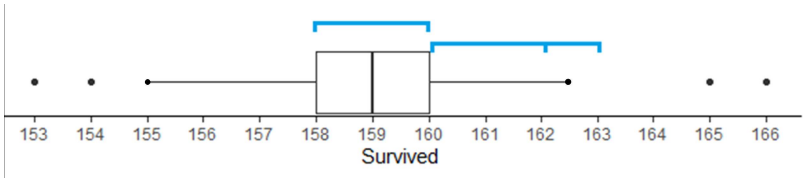


Boxplot w/whiskers

“Far away” pts should be outliers. What is “far away” enough?



Let’s use 1.5 times width of box as “too far away.”



Defining what an outlier is

Outlier Calculation

Recall: width of the box is IQR.

Outliers: pts more than 1.5 IQRs from nearest quartile.

- ▶ **Lower Outliers:** any pt below $LQ - 1.5(IQR)$.
- ▶ **Upper Outliers:** any pt above $UQ + 1.5(IQR)$.

Continuing our example:

Recall living sparrows: $LQ = 158$, $UQ = 160$.

Lower & upper outlier cut-offs?

$$IQR = 160 - 158 = 2.$$

$$1.5(IQR) = 1.5 \times 2 = 3.$$

Lower outlier is any pt below: $LQ - 1.5(IQR) : 158 - 3 = 155$.

Upper outlier is any pt above: $UQ + 1.5(IQR) : 160 + 3 = 163$.

Recall: Sparrows who Survived

153	154	154	155	156	156	157	157	158
158	158	158	159	159	159	159	159	159
160	160	160	160	160	160	160	160	160
161	161	161	161	162	163	165	166	



How many outliers are there?

Recall dead sparrows: $LQ = 161$, $UQ = 163.5$.

Lower & upper outlier cut-offs?

$IQR = 163.5 - 161 = 2.5$.

$1.5(IQR) = 2.5 \times 1.5 = 3.75$.

Lower outlier is any pt below: $LQ - 1.5(IQR) : 161 - 3.75 = 157.25$.

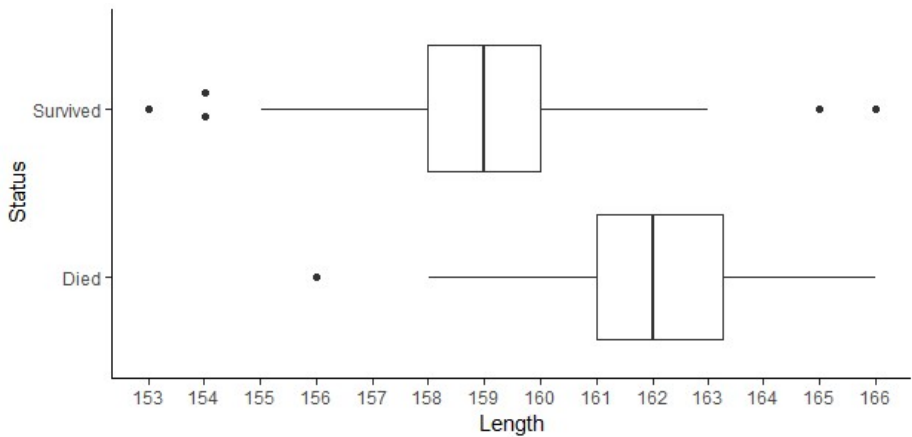
Upper outlier is any pt above: $UQ + 1.5(IQR) : 163.5 + 3.75 = 167.25$.

Recall: Sparrows who Died

156	158	160	160	160	161
161	161	161	161	162	162
162	162	162	162	163	163
164	164	165	165	166	166



How many outliers?



Modified Box Plots

Outliers are marked with a • .

! Whiskers go to last non-outlier.

Activities: 10-2, 10-3.

For 10-3:



bit.ly/introstatsdata

Data: IceCream

Applets: Dotplots

Convert grams to cups for icecream: [Inchcalculator.com/convert/gram-to-cup](https://inchcalculator.com/convert/gram-to-cup)

What did we learn?

- ♦ Five Number Summary (FNS)
- ♦ Box plot 
- ♦ Outliers 
- ♦ Modified Box Plot

