Types of Graphs

R

Lecture 4: Histograms

June 8, 2022

Switch 000 J Types



Instructor Switch

Course Administration

Good, Bad and Ugly

Variations of Graphs, Few Ch. 9

What is a Histogram?

R: ggplot and Histograms



McCall Needs to Depart – Please Send Best Wishes

- See Piazza email for McCall's message
- I strongly suspect she'd welcome words of support
- Everyone involved apologies for this mid-semester switch

• I want to adhere to everything McCall promised - so let me know

Switch

• New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?
- How did McCall do programming with you guys?

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?
- How did McCall do programming with you guys?
- Don't yet have access to google folder so get ready to share your screen for ${\sf G}/{\sf B}/{\sf U}$

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?
- How did McCall do programming with you guys?
- Don't yet have access to google folder so get ready to share your screen for ${\sf G}/{\sf B}/{\sf U}$
- I will return comments on policy brief proposals next week

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?
- How did McCall do programming with you guys?
- Don't yet have access to google folder so get ready to share your screen for ${\sf G}/{\sf B}/{\sf U}$
- I will return comments on policy brief proposals next week
- I will mean to post lecture notes, but I usually forget nag and I'll do it

- I want to adhere to everything McCall promised so let me know
- New webpage: go to www.leahbrooks.org \rightarrow teaching \rightarrow this semester
- Minor changes to schedule

- guest speaker replaced with in-class workshop
- switched classes 6 and 7 to make this work
- Did McCall say anything about fitting your presentations into one day?
- How did McCall do programming with you guys?
- Don't yet have access to google folder so get ready to share your screen for ${\sf G}/{\sf B}/{\sf U}$
- I will return comments on policy brief proposals next week
- I will mean to post lecture notes, but I usually forget nag and I'll do it
- Remaining concerns? we can return to this next class, as well



Introductions

- Name
- What program you're in
- What you do now
- What you want to do when you're done

Admin ●○

Types of 0

Course Administration

1. Reminder: Fully composed chart due June 15

- if there is something you want to do, but can't figure out how
- write it in words accompanying the graph
- I want to be sure you know what to do
- we can work on how to do it
- 2. Anything lingering?



Admin

Types of (

Turning in Tutorial Material

• What did McCall say to do?

Turning in Tutorial Material

- What did McCall say to do?
- write up your answers: one document that clearly lists question and answer
- don't make me (or anyone) have to read code
- give evidence that you have worked through the tutorial
 - R code
 - R output
- R Markdown instructions posted if curious resources tab



-

G/B/U

....

Next Week's Good Bad and Ugly

Find a histogram. Post by Friday at noon. Post the link on the google sheet.

| | Finder | Presenter |
|---|---------|-------------|
| 1 | Morre | Timberley |
| 2 | Spencer | Danielle G. |
| 3 | Evan L | Mary M. |
| 2 | Dustin | Nicole M. |



G/B/U

0.

This Week's Good Bad and Ugly

- You share
- Usually I'll stick into the slides

Which Graph for What Purpose?



Few: Three Basic Ways to Convey Information Graphically

- 1. Bars
- 2. Lines
- 3. Boxes for distributions

Types of Graphs 0000000

Bars



Admin 0

Types of Graphs

Lines

Population Turns Up After 2000





Source: https://www.r-graph-gallery.com/89-box-and-scatter-plot-with-ggplot2.html

| Relationship | Use | Avoid |
|--------------------|-----|-------|
| Nominal comparison | | |
| Time Series | | |
| Ranking | | |
| Part-to-whole | | |
| | | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-------|
| Nominal comparison | Bars, Points sparingly | |
| Time Series | | |
| Ranking | | |
| Part-to-whole | | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | | |
| Ranking | | |
| Part-to-whole | | |

.

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | Lines | |
| Ranking | | |
| Part-to-whole | | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | Lines | Bars falsely suggest independence |
| Ranking | | |
| Part-to-whole | | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | Lines | Bars falsely suggest independence |
| Ranking | Bars or Dots | |
| Part-to-whole | | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | Lines | Bars falsely suggest independence |
| Ranking | Bars or Dots | Not lines! |
| Part-to-whole | | |

| Rela | itionship | Use | Avoid |
|------|------------------|------------------------|-----------------------------------|
| Non | ninal comparison | Bars, Points sparingly | Bars starting above 0 |
| Tim | e Series | Lines | Bars falsely suggest independence |
| Ran | king | Bars or Dots | Not lines! |
| Part | -to-whole | Bars or stacked bars | |

| Relationship | Use | Avoid |
|--------------------|------------------------|-----------------------------------|
| Nominal comparison | Bars, Points sparingly | Bars starting above 0 |
| Time Series | Lines | Bars falsely suggest independence |
| Ranking | Bars or Dots | Not lines! |
| Part-to-whole | Bars or stacked bars | No pies! |

| Relationship | Use | Avoid |
|--------------|-----|-------|
| Distribution | | |
| Single | | |
| Multiple | | |
| Correlation | | |
| Geospatial | | |

.

| Relationship | Use | Avoid |
|--------------|---------------------------------------|-------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | | |
| Correlation | | |
| Geospatial | | |

.

| Relationship | Use | Avoid |
|--------------|---------------------------------------|-------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | Bars or Dots | |
| Correlation | | |
| Geospatial | | |

| Relationship | Use | Avoid |
|--------------|---------------------------------------|----------------------------------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | Bars or Dots | Two histograms together is hard! |
| Correlation | | |
| Geospatial | | |

.

| Relationship | Use | Avoid |
|--------------|---------------------------------------|----------------------------------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | Bars or Dots | Two histograms together is hard! |
| Correlation | Points or paired bars | |
| Geospatial | | |
Histogram

.

Types of Relationships You May Want to Show, 1 of 2

| Relationship | Use | Avoid |
|--------------|---------------------------------------|----------------------------------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | Bars or Dots | Two histograms together is hard! |
| Correlation | Points or paired bars | Rarely lines |
| Geospatial | | - |

Histogram

Types of Relationships You May Want to Show, 1 of 2

| Relationship | Use | Avoid |
|--------------|---------------------------------------|----------------------------------|
| Distribution | | |
| Single | Histogram, dot plot, or density curve | |
| Multiple | Bars or Dots | Two histograms together is hard! |
| Correlation | Points or paired bars | Rarely lines |
| Geospatial | Wait for maps! | |

| Switch | Admin | G/B/U | Types of Graphs | Histogram |
|--------|-------|-------|-----------------|---|
| 000 | 00 | 00 | 0000000 | •00000000000000000000000000000000000000 |

Histograms

What are non-graphical ways of describing the distribution of a variable?

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions. Ex.: Number of commuters by jurisdiction in DMV.

1. levels

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

- 1. levels
 - number of commuters by jurisdiction

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

Ex.: Number of commuters by jurisdiction in DMV.

- 1. levels
 - number of commuters by jurisdiction
- 2. shares

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

 $\label{eq:exact} \begin{array}{l} \mathsf{Ex.:} \ \mathsf{Number of \ commuters \ by \ jurisdiction} \\ \mathsf{in \ DMV}. \end{array}$

- 1. levels
 - number of commuters by jurisdiction
- 2. shares
 - share of commuters by jurisdiction

Histograms Show the Distribution of **One** Variable

What are non-graphical ways of describing the distribution of a variable?

- mean
- median
- mode
- variance
- percentiles

Two ways to think about distributions.

 $\label{eq:exact} \begin{array}{l} \mathsf{Ex.:} \ \mathsf{Number of commuters by jurisdiction} \\ \mathsf{in DMV.} \end{array}$

- 1. levels
 - number of commuters by jurisdiction
- 2. shares
 - share of commuters by jurisdiction

When do you want a histogram?

vitch

Types of 0

Histogram

- Take a variable
- Make bins by value
- Count the number of observations in each bin
- Plot bars with that number

Switch

A

Types of

Notionally, to Create a Histogram

Imaginary Income Data

| Person | Income |
|--------|--------|
| А | 4 |
| В | 11 |
| С | 12 |
| D | 3 |
| Е | 0 |

Switch

ch

Types of Graphs

| Imaginary Income Data | | Data Adding a Bin |
|-----------------------|--------|-----------------------------|
| Person | Income | Bin |
| А | 4 | 0, (or -4 to 0, to be even) |
| В | 11 | 1-5 |
| С | 12 | 5-10 |
| D | 3 | 10-15 |
| E | 0 | 15-20 |

Types of Graphs

Histogram R

| Imaginary Income Data | | a Adding a Bin | Bin Dataset | |
|-----------------------|--------|-----------------------------|-------------|-----|
| Person | Income | Bin | Bin | No. |
| A | 4 | 0, (or -4 to 0, to be even) | 0 | 1 |
| В | 11 | 1-5 | 1-5 | 2 |
| С | 12 | 5-10 | 5-10 | 0 |
| D | 3 | 10-15 | 10-15 | 2 |
| E | 0 | 15-20 | | |

Types of Graphs

Histogram R

| Imaginary Income Data | | Adding a Bin | Bin Dataset | | |
|-----------------------|--------|-----------------------------|-------------|----------------|--|
| Person | Income | Bin | Bin | No. | |
| A | 4 | 0, (or -4 to 0, to be even) | 0 | 1 | |
| В | 11 | 1-5 | 1-5 | 2 | |
| С | 12 | 5-10 | 5-10 | 0 | |
| D | 3 | 10-15 | 10-15 | 2 | |
| E | 0 | 15-20 | Graph th | Graph this one | |



B/U Typ

Key Features of Histograms

- Looks like a bar chart
- But! unlike a bar chart, histogram bars touch, to indicate continuity
- Which of Few's principles does this illustrate?

Switch 000 Admin

Types of 000000 Histogram

The Histogram Inventor

Karl Pearson (1857-1936) as a young man



ovitch

n G/

Types of (

The Histogram Inventor

Karl Pearson (1857-1936) as a young man



A big thinker

- father of mathematical statistics
- publishes first histogram, 1895
- fervent eugenicist
- early suffragist
- turned down knighthood due to socialist beliefs



Types of Graphs

Histogram Examples

- Income distribution
- As a guide on a map
- Income distribution for DC MSA

Switch 000 Adm

U Ty

Mulbrandon's Income Histogram



itch O /U Ty

Mulbrandon's Income Histogram



Switch

Admin

Types o

As a Map Legend

Drug poisoning deaths (2014)



From https://mathewkiang.com/2017/01/16/using-histogram-legend-choropleths/

G/B/U Types of Graphs His

Histogram

Density Curves: Smoothed Histograms

- Imagine many very thin bars
- This yields a curve
- Sometimes it is more helpful to draw the curve

Switch

G/B/U

Height: Note the Curves



From http://www.usablestats.com/lessons/normal

Income Distribution in the DC Metro Area Over Time

- goal is to show greater inequality and greater wealth in DC area, 2000 to 2018
- presentation and paper versions
- notice how they are different






















Types of Grap

Histogram

Income Distribution in the DC Metro Area Over Time



- print version
- was never satisfied with y axis
- light purple probably too light
- goal was to show 25th and 75th percentiles
- and change therein

| Switch Admin G/B/U Types of Graphs Flittogram K 000 00 00000000 000000000000000000000000000000000000 | Switch 000 |
|--|---------------|
|--|---------------|

R

To end of lecture

Today

- A. Heads-up: Bigger Data
- B. If-else recap
- C. Histograms
- D. Results by group: groupings and facets

A. Bigger Data

- > You need to work with more data than you can see in a window
- Today's tutorial has techniques to do this
- Look to summary statistics

A. Looking at crashes

| dim | (crash) | |
|-----|---------|--|
| dim | (crash | |

[1] 59777 44

table(crash\$Light)

##

| ## | DARK | UNKNOWN | LIGHTING | DARK LIGHTS ON | DARK NO LIGI |
|----|------|-------------|----------|----------------|--------------|
| ## | | | 660 | 13971 | 2: |
| ## | | | DAWN | DAYLIGHT | DI |
| ## | | | 1239 | 39305 | 1; |
| ## | | | N/A | OTHER | UNKN |
| ## | | | 497 | 143 | 4 |

| | look | at | the | total | size | of | the | dataset |
|--|------|----|-----|-------|------|----|-----|---------|
|--|------|----|-----|-------|------|----|-----|---------|

A. A Legible Version

| ## | # | A tibble: 9 x 2 | |
|----|---|-----------------------|-------------|
| ## | | Light | light_type |
| ## | | <fct></fct> | <int></int> |
| ## | 1 | DARK UNKNOWN LIGHTING | 660 |
| ## | 2 | DARK LIGHTS ON | 13971 |
| ## | 3 | DARK NO LIGHTS | 2158 |
| ## | 4 | DAWN | 1239 |
| ## | 5 | DAYLIGHT | 39305 |
| ## | 6 | DUSK | 1393 |
| ## | 7 | N/A | 497 |
| ## | 8 | OTHER | 143 |
| ## | 9 | UNKNOWN | 411 |

B. A Key Programming Command: ifelse()

B. An Example, 1 of 3

What if I want to know the century in which each building is built?

B. An Example, 2 of 3

Warning in Ops.factor(ex\$yb, 2000): '<' not meaningful for factors</pre>

B. An Example, 3 of 3

B. An Example, 3 of 3

table(ex\$c)

20th 21st ## 2 1

B. An Example, 3 of 3

table(ex\$c)

20th 21st ## 2 1

What could go wrong with programming like this?

B. Some rules of thumb for ifelse()

check your output!

B. Some rules of thumb for ifelse()

check your output!

- ▶ a test can include multiple conditions
- good idea to define all cases don't let a case be the residual

B. Some rules of thumb for ifelse()

- check your output!
- a test can include multiple conditions
- good idea to define all cases don't let a case be the residual
- you can nest ifelse() commands:

C. Histograms

We will use three new geoms this lecture

- geom_histogram()
- geom_density()
- geom_freqpoly()

C.1. How to create a histogram

```
Use
geom_histogram(data = [dataframe],
mapping = aes(x = [variable]))
```

- only need to list one variable
- histograms are univariate graphics
- geom_histogram() is best for a distribution with limited values

C.1. How to create a histogram

```
Use
geom_histogram(data = [dataframe],
mapping = aes(x = [variable]))
```

- only need to list one variable
- histograms are univariate graphics
- geom_histogram() is best for a distribution with limited values
- but not a categorical distribution, which should be a bar

C.2. Histogram options

- fill overall: outside aes, fill = [color]
- fill by group: inside aes, fill = [variable]
- bin width: bin_width = [unit span],
- by groups: inside aes, color = [grouping variable]

C.3. Approximating Continuous Distributions

For almost-continuous bins, use

geom_freqpoly()

For much more smoothing, use
geom_density()

C.4. Example

- take crash-level data from last class
- use group_by() and summarize() to make daily data
- count number of crashes by day

```
# add up total number of crashes by date
crash2 <- group_by(.data = crash, date2)
crash2 <- summarize(.data = crash2, tot.daily.crashes = n())
table(crash2$tot.daily.crashes)</pre>
```

##

| ## | 2 | 3 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| ## | 1 | 1 | 1 | 2 | 4 | 5 | 3 | 8 | 8 | 11 | 12 | 20 | 19 | 33 | 26 | 44 | 43 | 57 | 61 | 65 | 76 | 73 | 76 | 74 |
| ## | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 |
| ## | 80 | 76 | 85 | 89 | 66 | 59 | 61 | 55 | 59 | 48 | 38 | 27 | 37 | 39 | 18 | 25 | 18 | 11 | 14 | 8 | 8 | 6 | 6 | 5 |
| ## | 59 | 60 | 61 | 62 | 63 | 65 | 66 | 71 | 77 | | | | | | | | | | | | | | | |
| ## | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | | | | | | | | | | | | | | | |

Plot these data

Plot these data



D. Results by Group

```
# find the day of the week
crash2$day.of.week <- weekdays(x = crash2$date2)</pre>
```

check
table(crash2\$day.of.week)

##

| ## | Friday | Monday | Saturday | Sunday | Thursday | Tuesday | Wednesday |
|----|--------|--------|----------|--------|----------|---------|-----------|
| ## | 264 | 264 | 264 | 264 | 265 | 264 | 264 |

- you need a variable that indicates a group
- then plot distribution by group
- we'll use distribution of traffic accidents (x variable)
- by weekday (grouping variable)

By day of the week

By day of the week wd



-

By day of the week, better colors and thicker lines

By day of the week, better colors and thicker lines



```
By day of the week, facets
```

By day of the week, facets

wd



tot.daily.crashes



- Turn in Tutorial 4
- Turn in fully composed chart assignment to google folder
- Monmonier, How to Lie with Maps, Chapters 1 and 2
- Look at linked dot density map from Post