Data visualization tutorial—Instructor Notes*

Peter Carbonetto *University of Chicago*

Setup

- WiFi.
- Power outlets.
- Pace & questions (e.g., keyboard shortcuts).
- Reading what I type.
- Tutorial packet.
- TAs.
- Locating your team.
- Computer clutter.

Run `sessionInfo()`

Check the version of R that you are using:

```
sessionInfo()
```

Clear your R environment

The R environment is where all variables (and functions) are stored and accessed. You should start with an empty environment. Run this:

```
ls()
```

If this outputs names of objects, it means your environment is not empty, and you should restart R with a clean environment. Do either:

- `rm(list = ls())`.
- Or, in RStudio, **Session > Clear Workspace**.

Test plot

If the text is too small on the projector, use the following code to fix the size of the text for all subsequent plots:

```
theme_update(text = element_text(size = 30),
             axis.text = element_text(size = 30))
```

*This document is for the Data Visualization tutorial for the BSD qBio Bootcamp, MBL, 2019. **Current version**: July 22, 2019; **Corresponding author**: pcarbo@uchicago.edu.

Load data

Some examples of R code to check that the data were read correctly, and to inspect the table:

```
nrow(dat)
ncol(dat)
names(dat)
summary(dat)
object.size(dat)
```

Prepare the data for plotting

Code to filter rows that don't have the necessary data:

```
dat <- subset(dat,
              type == "Restaurant" &
              city == "CHICAGO" &
              !is.na(latitude) &
              !is.na(longitude) &
              !is.na(license))
```

Example code to extract data on pizza restaurants:

```
dat <- subset(dat,
  grepl("pizz",dba,ignore.case = TRUE) |
  grepl("pizz",aka,ignore.case = TRUE))
```

Example code to sort the rows of the table by year of inspection:

```
rows <- order(dat$year)
dat  <- dat[rows,]
```

Example code to create a new data frame containing one row per pizza location:

```
rows <- which(!duplicated(dat$license))
dat  <- dat[rows,]
```

Your first plot

Create a table containing the number of newly inspected restaurants by year:

```
counts <- table(dat$year)
```

Convert the "counts" to a data frame:

```
counts <- as.data.frame(counts)
names(counts) <- c("year","count")
```

Your second plot

```
library(ggplot2)
library(cowplot)
```

Create a plot showing a map of the pizza restaurants throughout the city:

```
a  <- aes(x = longitude,
          y = latitude)
p2 <- ggplot(dat,a)
g  <- geom_point()
p2 <- ggplot_add(g,p2)
```

Create a contour plot from the pizza locations, and overlay the points on top of it:

```
a  <- aes(x = longitude,
          y = latitude)
p2 <- ggplot(dat,a)
g  <- geom_density_2d()
p2 <- ggplot_add(g,p2)
g  <- geom_point()
p2 <- ggplot_add(g,p2)
```

Add colour corresponding to year:

```
a  <- aes(x     = longitude,
          y     = latitude,
          color = year)
p2 <- ggplot(dat,a)
g  <- geom_point()
p2 <- ggplot_add(g,p2)
```

Highlight more recent pizza restaurants in warmer (red) colours:

```
out <- scale_color_gradient2(low = "white",
         mid = "orange",high = "red",
         midpoint = 2014)
p2  <- ggplot_add(out,p2)
```

Save the combined plot as a PNG file:

```r
ggsave("plots.png",p12,dpi = 150)
```