

# Lab 8: Getting data into R and heatmaps

*Michael Lopez, Skidmore College*

## Important note 1

First, we'll open RStudio by going to <http://r.skidmore.edu/>.

## Important note 2

Open a new R Markdown file (File / New File / R Markdown...). You can create a basic name – Lab0, for example – and that'll set you up with a new file ready to go.

## Overview

### Today's goals

1. Data frame reminders
2. Accessing data using google sheets
3. Entering your data and seeing what it looks like

### What does your data look like?

- First row is a header
- First column is subject/unit identifier
- Avoid names, values or fields with blank spaces, otherwise each word will be interpreted as a separate variable
- If you want to concatenate words, inserting a . in between to words instead of a space
- Short names are preferred over longer names
- Avoid using names that contain symbols such as ?, \$, %, ^, &, \*, (, ), -, #, ?, <, >, /, |, , [ , ] , {, and };
- Delete any comments that you have made to avoid extra columns
- Missing values in your data set are indicated with NA

## Googlesheets

1. Here's a public link to a sample data set, stored on google sheets

([https://docs.google.com/spreadsheets/d/1wRAIt7W2mgaTqbdvutkZm\\_gOzJEUlaMcD7bIJj1UAgQ/edit#gid=1044528235](https://docs.google.com/spreadsheets/d/1wRAIt7W2mgaTqbdvutkZm_gOzJEUlaMcD7bIJj1UAgQ/edit#gid=1044528235))

2. If you haven't already, you can create your own google sheet by going to [sheets.google.com](https://sheets.google.com) and starting a new, blank spreadsheet
3. Next follow the link below to publish your Google Sheets to the web

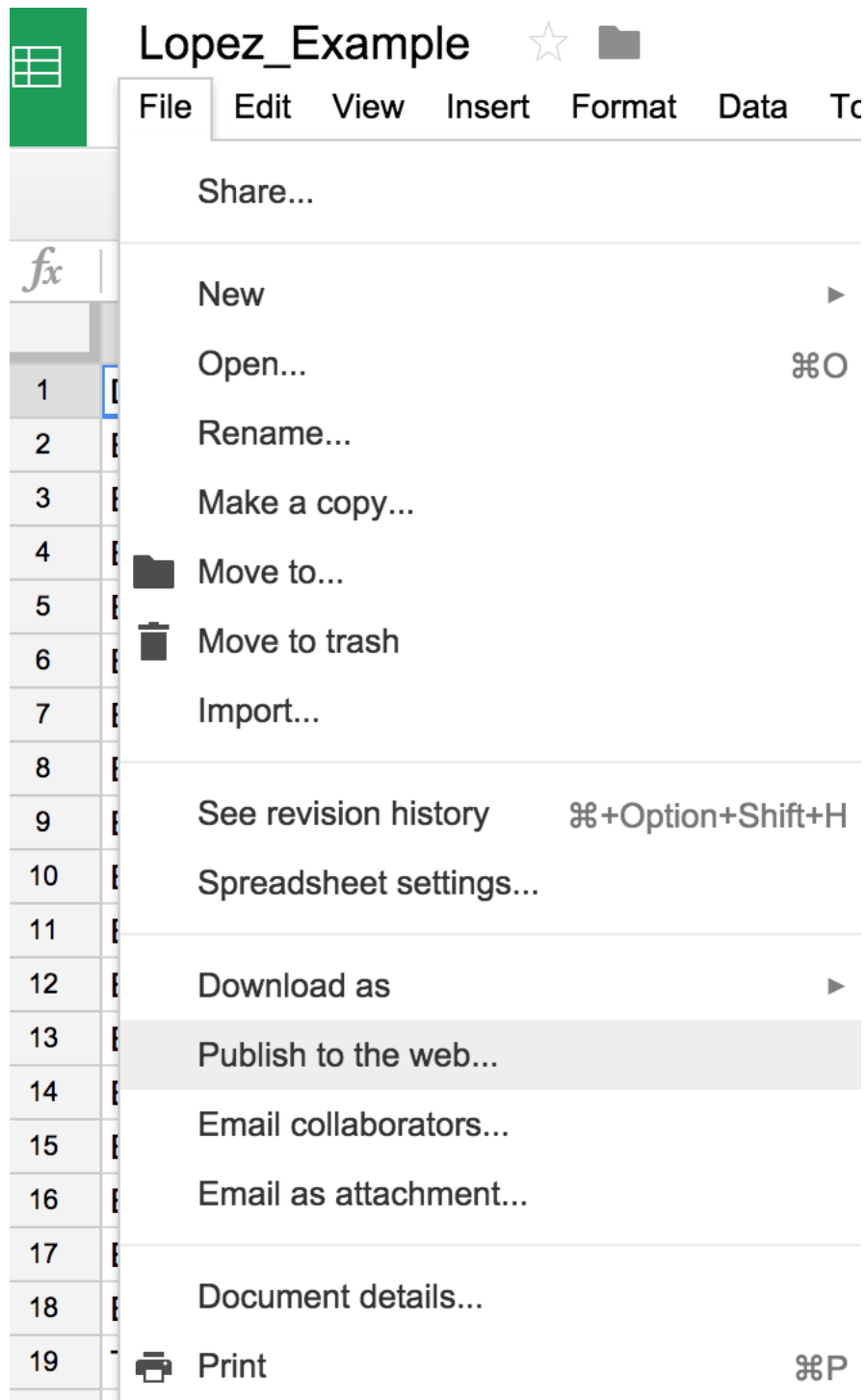


Figure 1: Publish to the web



## Publish to the web

This document is not published to the web.

Make your content visible to anyone by publishing it to the web. You can link to or embed your document. [Learn more](#)

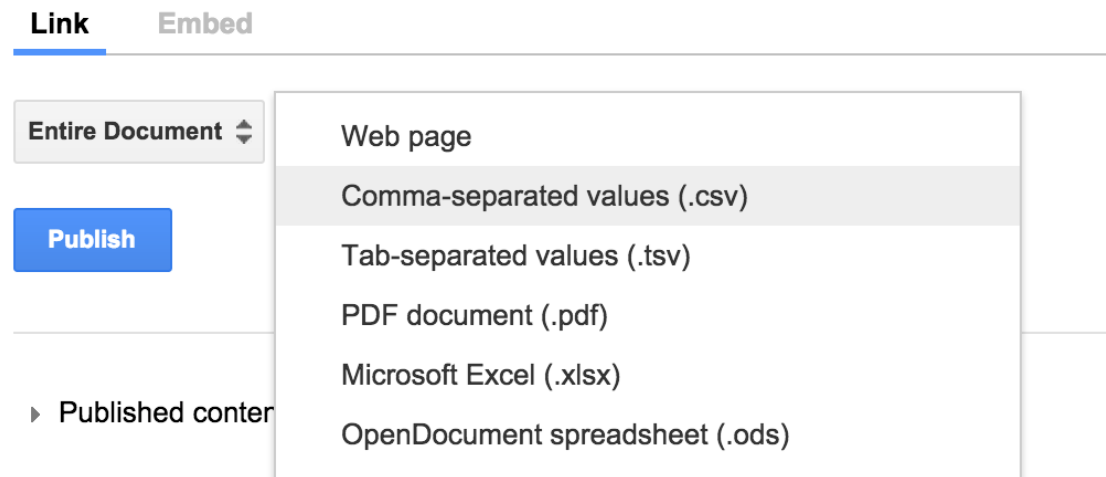


Figure 2: Make sure it's a csv

3. Next follow the link below to publish your Google Sheets to the web
4. Try it out!

First, try mine (as a reminder, make sure the URL is on one line)

```
library(dplyr)
url <- "https://docs.google.com/spreadsheets/
d/e/2PACX-1vSSVLvjUg5fgsyjmNrtnADtDunoS9n--
KeaTCu85yuSrkgrogIficedIHw07a14rhbDytoHyfQeFcgEu/pub?output=csv"
lopez.example <- read.csv(url)
head(lopez.example)
```

Second, try yours:

```
library(dplyr)
url <- "[Insert your url here]"
my.example <- read.csv(url)
head(my.example)
```

## Analyzing shot data

The data set in my Google spreadsheet looks at shots from the 2014-15 NBA season.

```
head(lopez.example)
ggplot(lopez.example, aes(x=LOC_X, y=LOC_Y)) +
  geom_point(aes(colour = EVENT_TYPE))
```

1. What do we notice about the distributions of shots in NBA games?
2. Provide a range of the LOC\_X and LOC\_Y variables. Is the NBA court only one side?
3. Summarize the SHOT\_ZONE\_BASIC variable below.

```
ggplot(lopez.example, aes(x=LOC_X, y=LOC_Y)) +
  geom_point(aes(colour = SHOT_ZONE_BASIC, shape = EVENT_TYPE)) +
  xlim(-250, 250) +
  ylim(-50, 420)
```

4. The following code sets up a variable PTS. What do the variables pt.value and PTS represent?

```
lopez.example <- lopez.example %>%
  mutate(pt.value = ifelse(SHOT_TYPE == "2PT Field Goal", 2, 3),
         PTS = pt.value*SHOT_MADE_FLAG)
head(lopez.example, 10)
```

5. Identify which of the shot zones (SHOT\_ZONE\_BASIC) yields the highest pts - per - shot.
6. What is done in the following chart?

```
library(tidyverse)
ggplot(lopez.example, aes(x=LOC_X, y=LOC_Y)) +
  stat_binhex(bins = 25, colour = "gray", alpha = 0.7) +
  scale_fill_gradientn(colours = c("yellow", "orange", "red")) +
  guides(alpha = FALSE, size = FALSE) +
  xlim(250, -250) +
  ylim(-52, 418) +
  geom_rug(alpha = 0.2) +
  coord_fixed()
```

7. Use the help commands or Google to identify what geom\_rug and coord\_fixed() do.
8. Facet the above graph by PLAYER\_NAME and compare each player.