Introduction to R

Mica Jorgenson John Fink

What is R?

- A programming language that lends itself well to statistical analysis analysis, text analysis, and text processing

Why R?

- Open source
- Used widely in Academia
- Data wrangling lots of packages available for dealing with "messy" data
- Visualization
- Machine learning (trainable)



Workshop Structure

- 1. Swirl (Do as many tutorials as you can in 1h!)
- 2. Break
- **3.** Applied R (Play with City of Hamilton Film Permit Data)



swirl - swirlstats.com

- 1. Open R Studio
- 2. Install swirl
 - > install.packages("swirl")
- 3. Start swirl
 - > library("swirl")
 - > swirl()



swirl commands

You can exit swirl and return to the R prompt (>) at any time by pressing the Esc key. If you are | already at the prompt, type bye() to exit and save your progress. When you exit properly, you'll | see a short message letting you know you've done so.

| When you are at the R prompt (>):

- | -- Typing skip() allows you to skip the current question.
- | -- Typing play() lets you experiment with R on your own; swirl will ignore what you do...
- | -- UNTIL you type nxt() which will regain swirl's attention.
- | -- Typing bye() causes swirl to exit. Your progress will be saved.
- | -- Typing main() returns you to swirl's main menu.
- | -- Typing info() displays these options again.

Experiment time! City of Hamilton Film Permits

https://bit.ly/2FqgAp2

Hamilton					Find	Find Q My Data Sign In		
Overview	Data							
Film Permits					★ Favorite - Download - APIs -			
Custom License List of film per Attribute	■ 10/19/201 mits issued by €S	Tabular Dataset the City of Hamilton	■ 497 Rows			spreadsheet About Cig of Hamilton - Web GIS Framework Open Date Shared Bp: coh_opendata Buta Source spatialok/dism/Jamilton.ca		
L CATEGORY Text	LEND_DATE	PERMIT_NUMBER Text	PRODUCTION_COMPANY	PRODUCTION_NAME	LAND START_DATE	View Metadata Create Webmap Create a Story Map		
MARD Text	OW MORE 1 Attribute							
Related Data								
	Building Permits - Construction Value by Permit Type Annual summary data of building construction permits issued by the Cry of Hermiten's Building Division as regulated by the TARLE TARLE					Nermits - Construction Value by Year as of total encuat building permits issued by the City I Building Division encore 1998 as regulated by the		
	To	bular Dataset			Tabular Date	aset		
				South and the second se				



Getting Hamilton Film Data into R

Open the data in R and assign it to a variable:

```
> film <- read.csv(file.path("C:",
"/Users/micaa/Downloads/Film_Permits.csv"))
```

> summary(film)

```
OR:
```

```
film <- read.csv(file="Film_Permits.csv")</pre>
```

If you're having trouble: check what directory R is working in using getwd ()

Make sure the file is in your working directory.

Asking basic questions

How many film permits?

>str(film)

How many were documentaries vs features vs Commercial vs. Series, etc?
Category <- film[,5]</pre>

summary(Category)



Asking Basic Questions con't

How many film permits were issued in Ward 1 vs. Ward 13 vs. Ward 1-15?

> summary(film[,6])

How many permits were issued in each Hamilton ward, on average?

> mean(table(film\$ward))



Make a pie chart showing the number of permits issued in each ward.

> pie(table(film\$ward))

Try a bar graph.

> barplot(table(film\$ward))

> wards <- table(factor(films\$WARD, levels =
c("1","2","3","4","5","6","7","8","9","10","11","12","13","1
4","15","1 to 15")))</pre>

> barplot(wards)

> barplot(wards, ylim = c(0, 100))





The Programming Historian (has a bunch of R tutorials. Start with this one): <u>https://programminghistorian.org/en/lessons/basic-text-processing-in-r</u>

R for Data Science (working with tidyverse):

https://r4ds.had.co.nz/

Seeing Theory (an intro to stats):

https://seeing-theory.brown.edu/

