



BUILDING  
TOMORROW  
TODAY



# for Data Science

An introduction to R statistical computing for the growing trend of data science,

R uses,  
capabilities,  
community,  
resources &  
a demonstration in RStudio

Open to data analytic enthusiasts

247/293,  
University of Botswana,  
5pm - 6pm,  
3rd Thursday March 2022


 [rladies/meetup\\_presentations\\_gaborone](https://github.com/rladies/meetup_presentations_gaborone)

 @RladiesGaborone



# Overview

What is data science

What is 

How data science is used

R for data science

R for research

R vs Python

RStudio

Where R is used

RStudio Demonstration



# Acknowledgments



# Who we R



**Ontiretse Ishmael**

R-Ladies co-organizer  
Teaching Assistant since 2017,  
Msc in Computer Science  
Department of Computer Science,  
University of Botswana

 [@RladiesGaborone](#) / [@ontizy](#)



**Simisani Ndaba**

R-Ladies Gaborone co-organiser,  
Teaching Assistant since 2016,  
Department of Computer Science,  
University of Botswana Msc in  
Computer Information Systems

[@RladiesGaborone](#) / [@simisani10](#)



[simisani.ndaba@013gmail.com](mailto:simisani.ndaba@013gmail.com)



## Past Meetup Events




**R Data Analysis in Medical Science Part 2: Practical Demonstration in RStudio with Mouneem Essabbar, PhD**  
 @mouneem



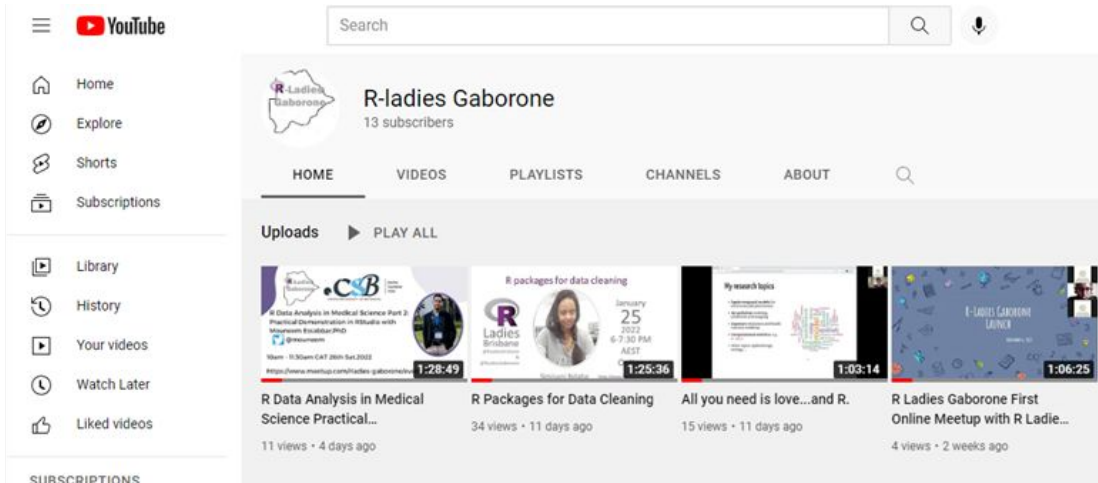
10am - 11:30am CAT 26th Sat, 2022  
<https://www.meetup.com/rladies-gaborone/events/283962337/>




**R packages for data cleaning**

January 25 2022  
 6-7:30 PM AEST  
 Online

Simisani Ndaba  
<https://www.meetup.com/rladies-brisbane>



YouTube channel page for R-ladies Gaborone (13 subscribers). The page shows a list of uploads:

- R Data Analysis in Medical Science Practical...** (11 views • 4 days ago)
- R Packages for Data Cleaning** (34 views • 11 days ago)
- All you need is love... and R.** (15 views • 11 days ago)
- R Ladies Gaborone First Online Meetup with R Ladies...** (4 views • 2 weeks ago)



**"ALL YOU NEED IS LOVE... AND R!"**

ACADEMIC (TEACHING AND RESEARCH) ACTIVITIES USED IN R WITH MICHELA CAMELETTI FROM THE UNIVERSITY OF BERGAMO

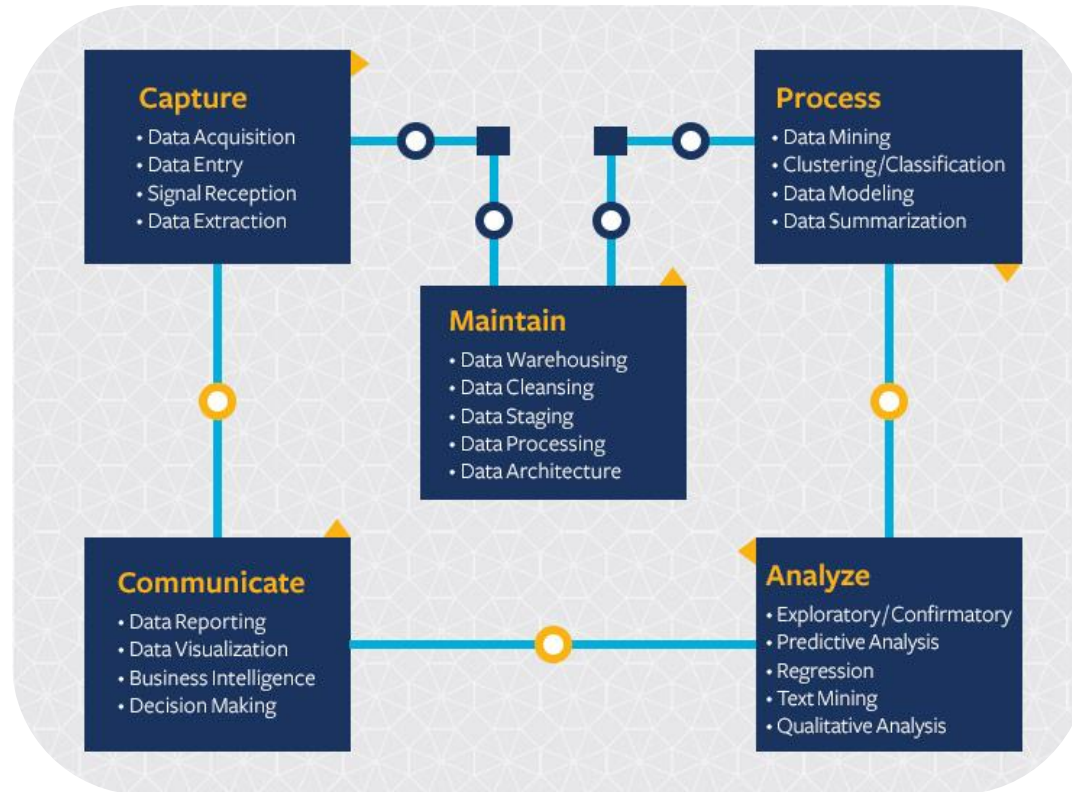
**SAT DEC 11, 2021 10AM - 11AM CAT**

[HTTPS://WWW.MEETUP.COM/RLADIES-GABORONE/EVENTS/281989529/](https://www.meetup.com/rladies-gaborone/events/281989529/)





# Data Science





# How Data Science is used



2019 Liverpool F.C.

Research team built;  
Proprietary  
model(probability+risk analysis)  
Overall chance of winning=  
runs,+goal attempts passes

Still used for General Strategy  
and Recruitment



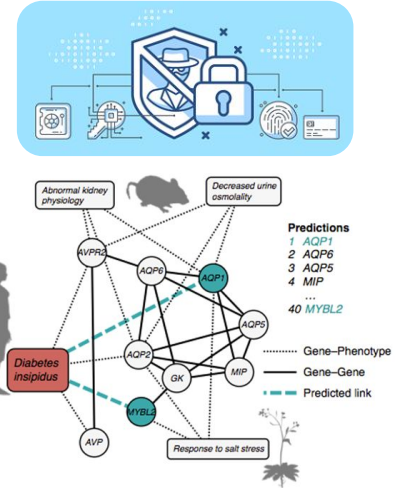
Prioritise=  
active users +

users near  
each other +

users who  
seem like each  
other's "types"  
based on their  
swiping  
history.



Plots from  
data  
coordinates



# What is



R is a language and environment for statistical computing and graphics.

The primary uses of R is and will always be, statistic, visualization, and machine learning.

R is developed by academics and scientist.

R possesses an extensive catalog of statistical and graphical methods. It includes;



machine learning algorithms, deep learning linear regression,  
time series, statistical inference, data mining

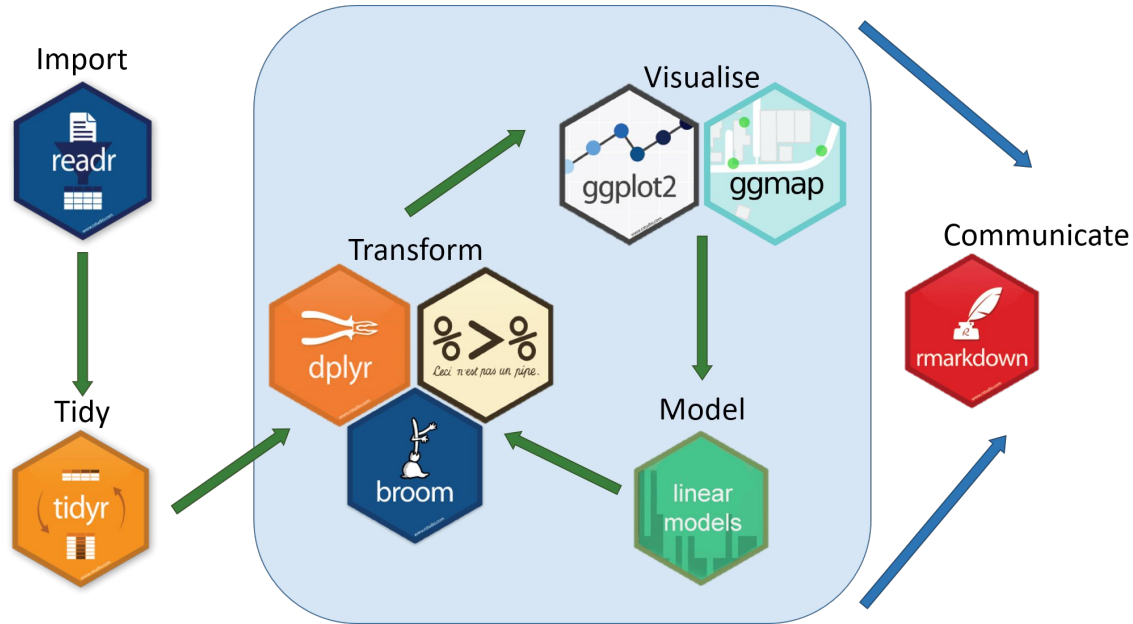


<https://www.r-project.org/about.html>





# R for Data Science



*R Data science workflow, (Hadley Wickham) R for Data Science*

# R for Research

The most important task in data science research is the way you deal with the data: import, clean, prep, feature engineering, feature selection.

R part of The Carpentries, teaches foundational coding and data science skills to researchers

F A I R data principles - Findable Accessible Interoperability Reproducible



# R vs Python

## Python

Open Source  
data manipulation and  
graphing  
Machine learning interfaces  
more general programming  
language

web development, and  
software development

## R

Open Source  
data manipulation and  
graphing  
Machine learning interfaces

typically used in statistical  
computing

R

```
library(purrr)
library(dplyr)
ba %>%
  select_if(is.numeric) %>%
  map_dbl(mean, na.rm = TRUE)
```

```
player NA
os NA
ge 26.5093555093555
ref_team_id NA
output truncated]
```

Python

```
nba.mean()
```

```
age 26.509356
53.253638
s 25.571726
output truncated]
/code>
```

Finding Averages for Each Statistic

R

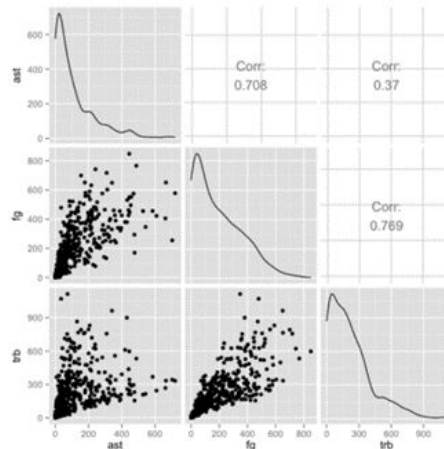
```
library(randomForest)
redictorColumns <- c("age", "mp", "fg", "trb", "stl", "blk")
f <- randomForest(train[predictorColumns], train$ast, ntree=100)
redictions <- predict(f, test[predictorColumns])
```

Python

```
from sklearn.ensemble import RandomForestRegressor
redictor_columns = ["age", "mp", "fg", "trb", "stl", "blk"]
f = RandomForestRegressor(n_estimators=100, min_samples_leaf=3)
f.fit(train[predictor_columns], train["ast"])
redictions = f.predict(test[predictor_columns])
```

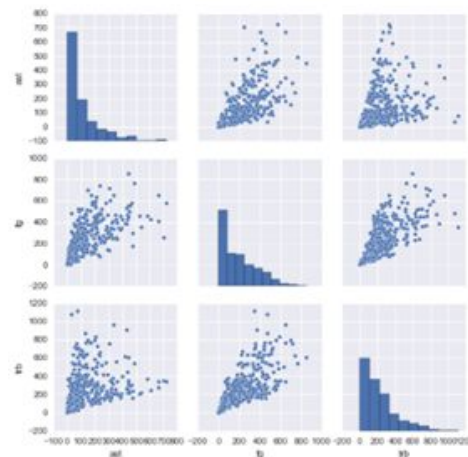
R

```
library(GGally)
ba %>%
  elect(ast, fg, trb) %>%
  gpairs()
```



Python

```
import seaborn as sns
span class="token keyword">import matplotlib.pyplot as plt
ns.pairplot(nba[["ast", "fg", "trb"]])
lt.show()
```



Making Pairwise Scatterplots

Fit a random forest model

```

Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Terminal Jobs
i12 ~ /
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (c) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

This free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> R package loaded from ~/.RData

> i("AirPassengers")
> i(airquality$Ozone)

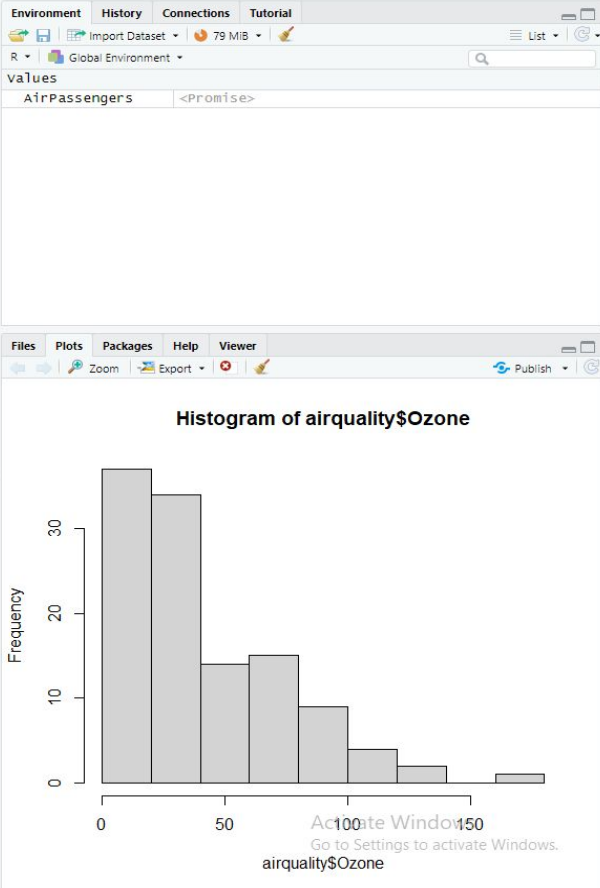
```

airquality

Filter

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.4	67	5	1
36	118	8.0	72	5	2
12	149	12.6	74	5	3
18	313	11.5	62	5	4
NA	NA	14.3	56	5	5
28	NA	14.9	66	5	6
23	299	8.6	65	5	7
19	99	13.8	59	5	8
8	19	20.1	61	5	9
NA	194	8.6	69	5	10
7	NA	6.9	74	5	11
16	256	9.7	69	5	12
11	290	9.2	66	5	13
14	274	10.9	68	5	14

1 to 14 of 153 entries, 6 total columns



RStudio Cloud

rstudio.cloud/project/3600336

Your Workspace / Untitled Project + Click to name your project

RAM

simisani ndaba

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

R 4.1.2

simi.R x AirPassengers x

Show Attributes

Name	Type	Value
AirPassengers	double [144] (S3: ts)	112 118 132 129 121 135 ...
(attributes)	list [2]	List of length 2

Environment History Connections Tutorial

R - Global Environment

Values

AirPassengers Time-Series [1:144] from 1949 to 1961: ...

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Name	Size	Modified
..		
.Rhistory	0 B	Feb 16, 2022, 8:39 AM
project.Rproj	205 B	Feb 16, 2022, 8:48 AM
simi.R	70 B	Feb 16, 2022, 8:49 AM

(No selection)

Console Terminal Jobs

R 4.1.2 - /cloud/project/

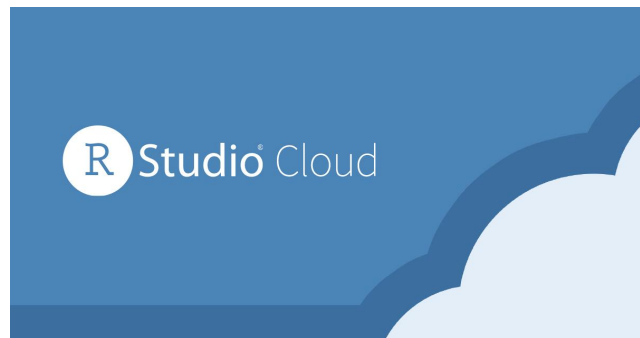
Type 'license()' or 'licence()' for distribution details.

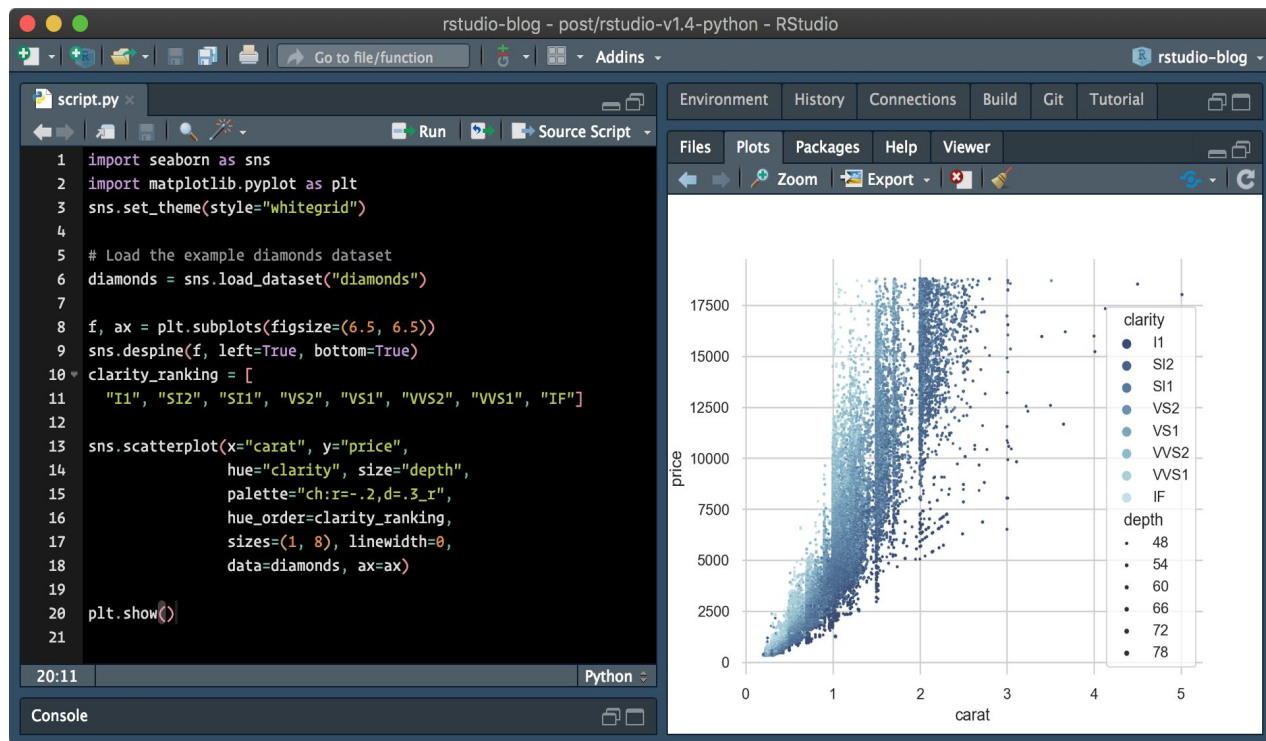
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

```
> getwd()
[1] "/cloud/project"
> getwd()
[1] "/cloud/project"
> data("AirPassengers")
> head(AirPassengers)
[1] 112 118 132 129 121 135
> View(AirPassengers)
> |
```

Activate Windows  
Go to Settings to activate Windows.





`install.packages("reticulate")`





# Where R is used in Botswana

## Stanbic Bank

Develop custom data models and algorithms to apply datasets using statistical computer language to build machine learning models

## FnB

Data analysis



## Old Mutual

Actuarial Support



## Mascom

Data analysis



## BIHL

Machine Learning models



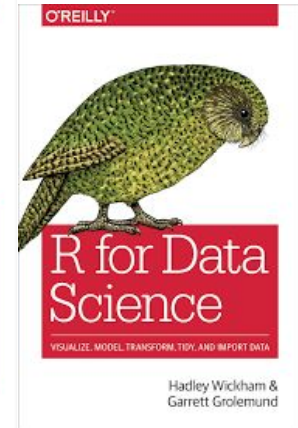
# R Community & Resources



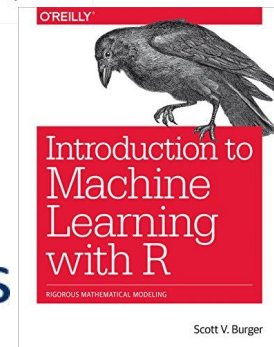
consortium



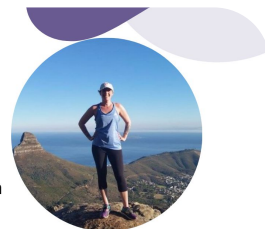
Garrett Grolmund  
reword by Hadley Wickham



Hadley Wickham &  
Garrett Grolmund



Scott V. Burger



Using R as an epidemiologist:  
assessment of real-world data with  
Brianna Lindsay, PhD

<https://www.meetup.com/rladies-gaborone/events/283714367/>

10am (CAT) Saturday 12th March, 2022  
R-Ladies Dammam and R-Ladies Gaborone



An Introduction to R Shiny with Mohamed EL  
Fodil Ihaddaden, Ph.D, Shiny developer

19:00pm 25th Friday March, 2022

<https://www.meetup.com/rladies-gaborone/events/283964083/>



How To Time Travel with your Code:  
foundations of version control, useful git commands and connections with RStudio  
with Gracielle Higino, a computational ecologist

10am - 11:30am CAT Saturday 30th April 2022

<https://www.meetup.com/rladies-gaborone/events/284115113/>

# Future Events

**Next Presentation on:**

**10th Thursday March, 5pm - 6pm**  
**293/247**



# RStudio Demonstration

Example;

```
install.packages("ggplot2")
```

```
install.packages("dplyr")
```

```
install.packages("ggmap")
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(ggmap2)
```

# Thank you



[@RladiesGaborone](#)



[R-ladies Gaborone](#)



**GitHub**

[meetup-presentations-Gaborone](#)



[R-Ladies Gaborone](#)

