West Michigan Users Group

Slides:
bit.ly/BDI_2018

Agenda: 1st Session

BIG DATA IGNITE

What is R? What is WMRUG?

Announcements

Upcoming conferences, local R jobs, Current Events

What is R?

Paul Egeler, Spectrum Health

R for Report Automation

Aaron Clark, Advanced Radiology Services

Slides:
bit.ly/BDI_2018

Agenda: 2nd & 3rd Session

When Less is More: Picking your R (starter) Toolkit

Randall Pruim, Calvin College

9:20am - Grand Gallery B

Statistical Machine Learning with R

David Zeitler, Grand Valley State University

10:25am - Grand Gallery B

What is R?

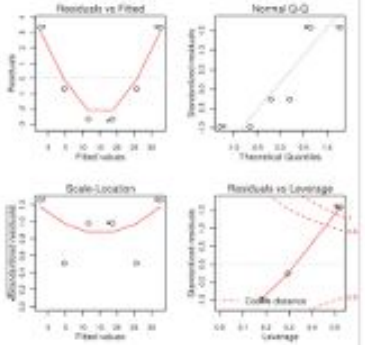
BIG DATA IGNITE

what is R?

All Videos News Shopping Images More Settings Tools

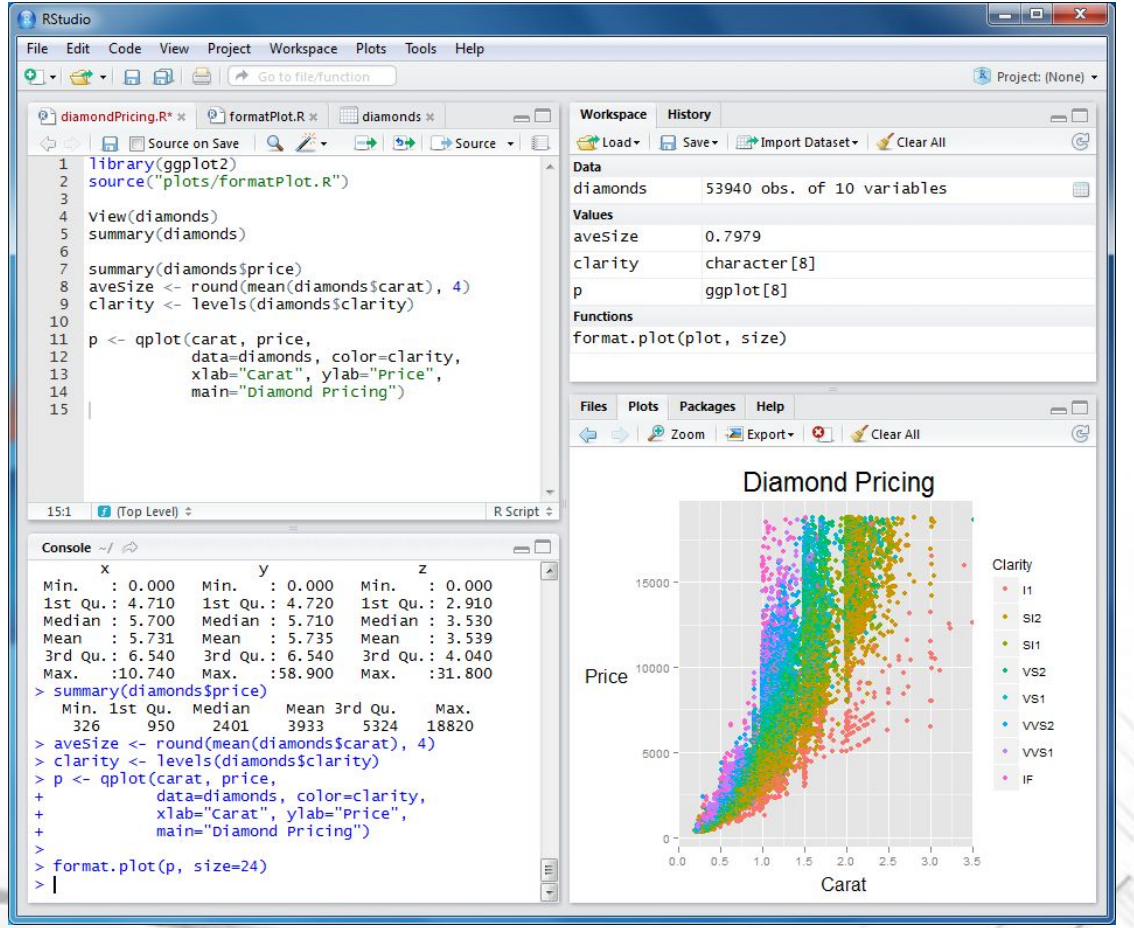
About 1,630,000,000 results (0.71 seconds)

R is an open source programming language and software environment for statistical computing and graphics that is supported by the **R** Foundation for Statistical Computing. The **R** language is widely used among statisticians and data miners for developing statistical software and data analysis.



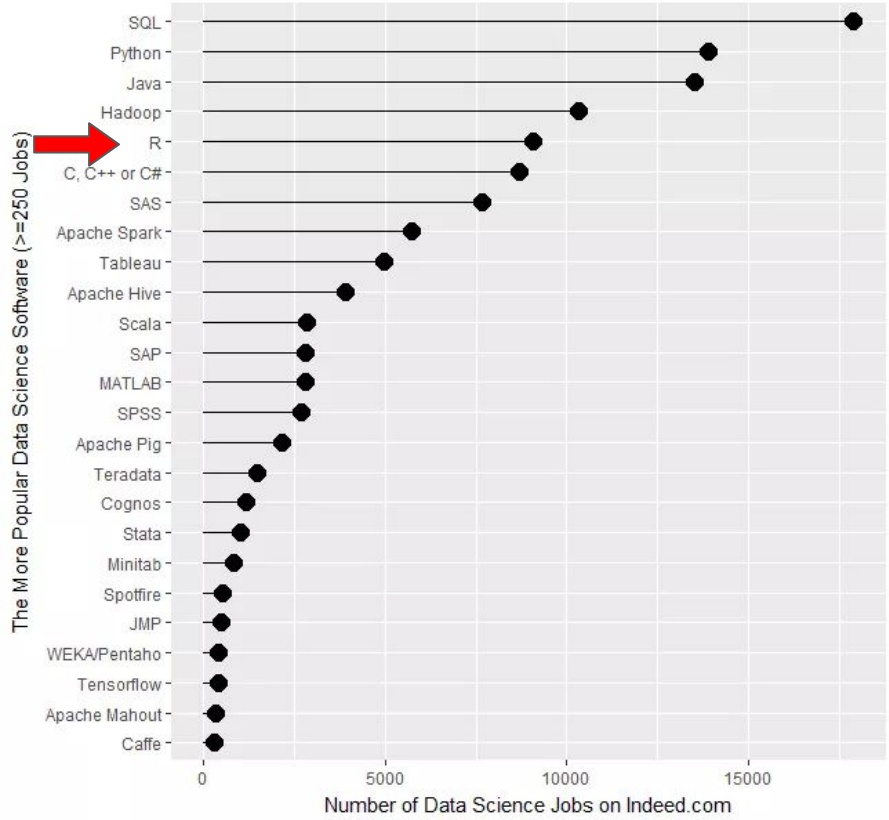
R (programming language) - Wikipedia
https://en.wikipedia.org/wiki/R_(programming_language)

What is R?


RStudio
File Edit Code View Project Workspace Plots Tools Help
diamondPricing.R* formatPlot.R diamonds
Source on Save Source
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12 data=diamonds, color=clarity,
13 xlab="Carat", ylab="Price",
14 main="Diamond Pricing")
15
15:1 (Top Level) R Script
Console ~/
x y z
Min. : 0.000 Min. : 0.000 Min. : 0.000
1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910
Median : 5.700 Median : 5.710 Median : 3.530
Mean : 5.731 Mean : 5.735 Mean : 3.539
3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040
Max. :10.740 Max. :58.900 Max. :31.800
> summary(diamonds$price)
Min. 1st Qu. Median Mean 3rd Qu. Max.
326 950 2401 3933 5324 18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+ data=diamonds, color=clarity,
+ xlab="Carat", ylab="Price",
+ main="Diamond Pricing")
>
> format.plot(p, size=24)
>
Workspace History
Load Save Import Dataset Clear All
Data
diamonds 53940 obs. of 10 variables
Values
aveSize 0.7979
clarity character[8]
p ggplot[8]
Functions
format.plot(plot, size)
Files Plots Packages Help
Zoom Export Clear All
Diamond Pricing
Price
Carat
Clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
0
5000
10000
15000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

What is R?

Data Science Jobs by Language/Software (Feb 2017)



http://r4stats.com/articles/popularity/

IEEE 2018 Top Enterprise Programming Languages: Derived from search, social media, and job listing trends

Language Rank	Types	Spectrum Ranking
1. Python	Web, Enterprise, Embedded	100.0
2. C++	Mobile, Enterprise, Embedded	99.7
3. Java	Web, Mobile, Enterprise	97.5
4. C	Mobile, Enterprise, Embedded	96.7
5. C#	Web, Mobile, Enterprise	89.4
6. PHP	Web	84.9
7. R	Enterprise	82.9
8. JavaScript	Web, Mobile	82.6
9. Go	Web, Enterprise	76.4
10. Assembly	Embedded	74.1

Language Types (click to hide)

Web | Mobile | Enterprise | Embedded

https://spectrum.ieee.org/static/interactive-the-top-programming-languages-2018

What is WMRUG?

<-'West Michigan R Users Group'

BIG DATA IGNITE

A Community to

- **Learn**
- **Share**
- **Innovate**
- **Troubleshoot**
- **Network**

A Mini BDI: For R-Users, by R-users!
Meet Monthly on 3rd Tuesdays
Usually @ GVSU's Downtown Campus

bit.ly/BDI_2018

Meetup

in

Call for Presenters!

Lightning Talks:

- 10 min talk with 5 minutes for questions
- Great opportunity for students

R Use-cases

- 30-50 min talk with time for questions

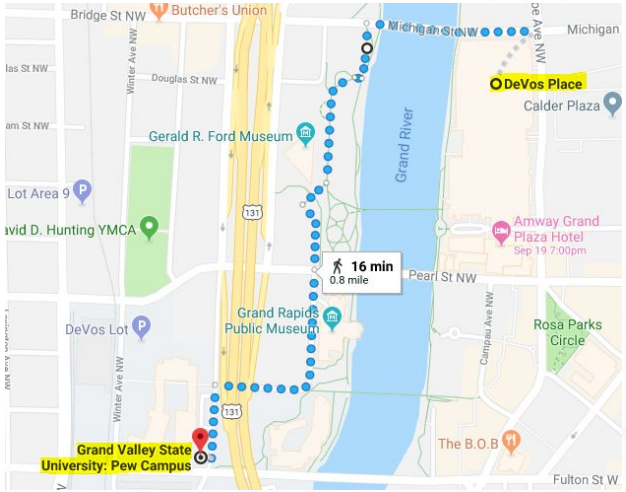
In Either case...

- Get R-swag
- Add presentation to Resume/CV
- Share your knowledge!

Next Meeting: Tues, Oct 16th

R for the Full Analytics Stack
Paul Egeler, Spectrum Health

6p - 8p
DEV 121E
Pew Campus
GVSU



Announcements

Upcoming Conferences/Talks with R Topics

Global Data Science Conference: Sep 25 - 27
Boston, MA
http://globalbigdataconference.com/

Open Data Science Conference: Oct 31 - Nov 3
San Francisco, CA
https://odsc.com/

#ODSC

Data Science Pop-up: Oct 30
Chicago, IL
https://popup.dominodatalab.com/

DOMINO DATA SCIENCE POPUP

Rstudio Conf: Jan 15 - 18
Austin, TX
https://www.rstudio.com/conference/

rstudio::conf
from RStudio

Announcements

New Remote R Jobs

In the last month… via r-users.com

Part-Time **Big Data Discovery Automation in Digital Health (5-10 hours per week)** Anywhere 7 Sep 2018
MedStar Institutes for Innovation – Posted by Praxiteles

Full-Time **R&D Database Developer** 10 Aug 2018
Crescendo Technology Ltd – Posted by Crescendo – Anywhere

Full-Time **Algorithm Developer** 10 Aug 2018
Crescendo Technology Ltd – Posted by Crescendo – Anywhere

REMOTE Senior Data Scientist- Supply Chain Forecasting
Surge - No office location
Remote
r python sql machine-learning

Predictive Modeling & Data Analytics Lead
DW Simpson Global Actuarial Recruitment
United States

Announcements

BIG DATA IGNITE

New Local R Jobs

In the last month… via r-users.com

Senior Software Developer - new
Initech Global
Grand Rapids, MI 49546
Needs experience in object-oriented programming on Python, R, java or .Net

Research Associate - Cancer Metabolism
Van Andel Institute ★★★½☆ 8 reviews
Grand Rapids, MI
R), and experience with the analysis of large data sets (i.e. Research Associate –

Bioinformatics Research Scientist
Van Andel Institute ★★★½☆ 8 reviews
Grand Rapids, MI
Experience in R (Required) and at least one other programming/scripting language

Computational Biologist - Epigenetics in Metabolism
Van Andel Institute ★★★½☆ 8 reviews - Grand Rapids, MI

Consulting Advanced Analytics Analyst - new
Steelcase ★★★★☆ 436 reviews
Grand Rapids, MI 49508

Research Associate-Fixed Term - new
Michigan State University ★★★★½ 543 reviews
Grand Rapids, MI
The ideal candidate should have experience with eye-tracking methodology and analyses, programming knowledge, and experience using R, MATLAB, or Python

Senior Business Intell Analyst (Supply Chain)
Spectrum Health ★★★★☆ 602 reviews - Walker, MI

Computational biologist - peds oncology program
Spectrum Health ★★★★☆ 602 reviews
Grand Rapids, MI
Programming experience in R and Unix. Experience - Bioinformatics, especially DNA sequence analysis, R programming, and laboratory information management...

Biostatistician Lead Peds Oncology Research Program
Spectrum Health ★★★★☆ 602 reviews
Grand Rapids, MI

Pricing Data Analyst
Meijer ★★★½☆ 6,648 reviews - Grand Rapids, MI

BIG DATA IGNITE

1st Lightning Talk

What is R?

-Paul Egeler, Spectrum Health

Overview

R is a Statistical Programming Language

Free and Open Source Software (FOSS)

Highly Extensible

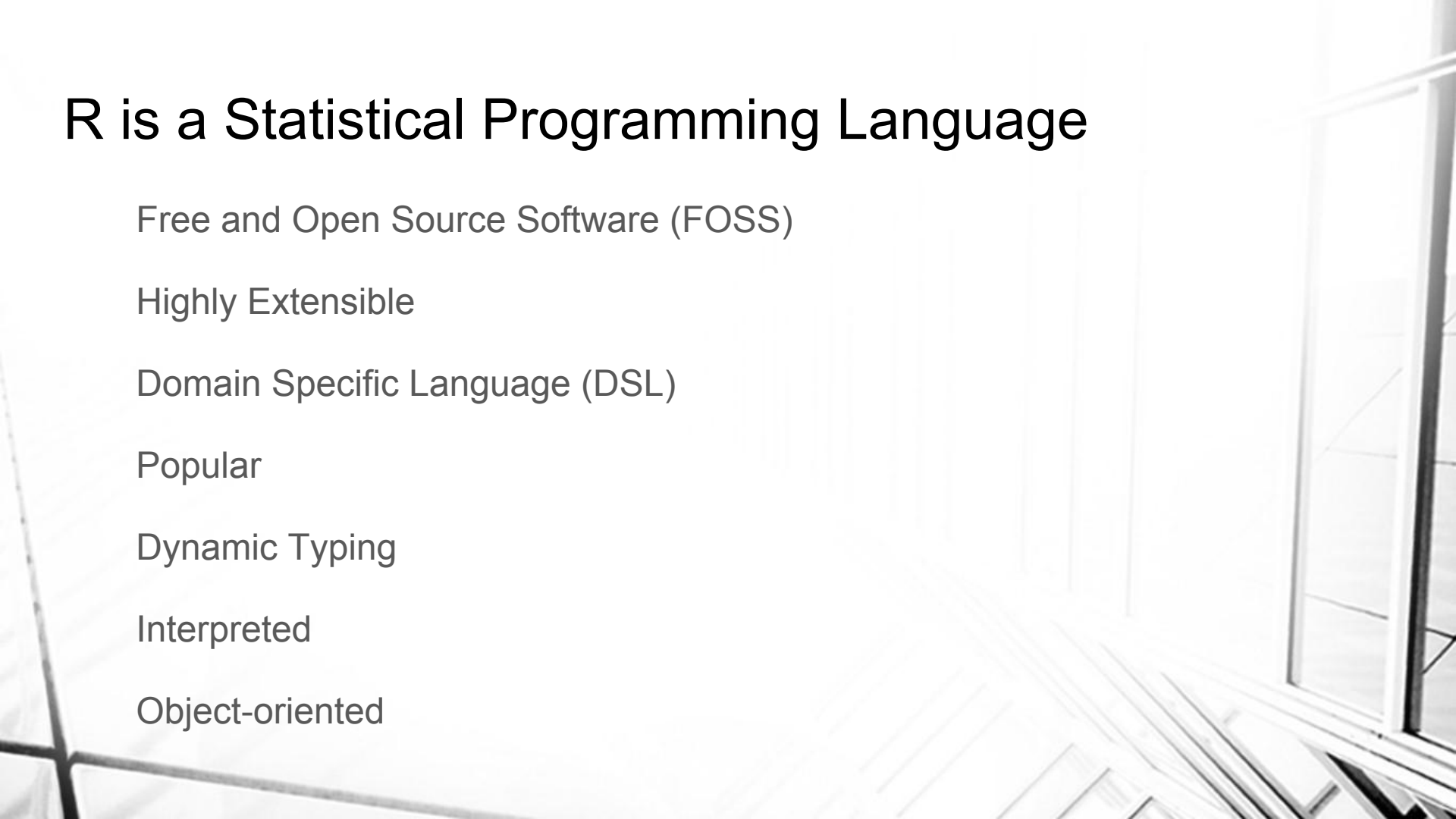
Domain Specific Language (DSL)

Popular

Dynamic Typing

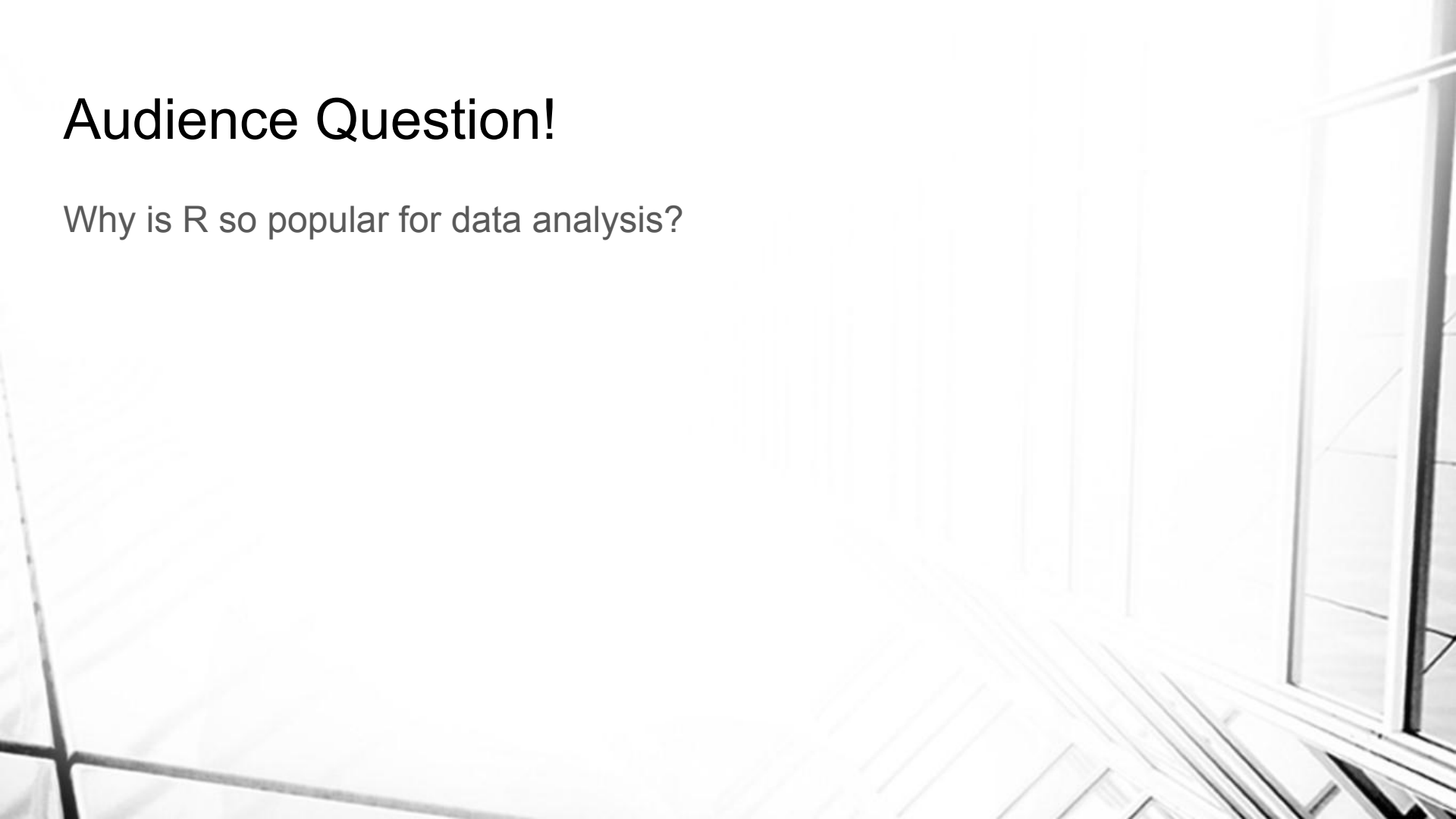
Interpreted

Object-oriented



Audience Question!

Why is R so popular for data analysis?



How do you interact with R?

Command Line Interpreter

Batch

Inside other analytics and BI tools

Integrated Development Environment (IDE)

- **RStudio**
- ESS
- VS Code
- VIM
- Eclipse
- Sublime
- Many others!

Use Cases

- Statistics
- Data Mining / Machine Learning
- Plots and Visualization
- Bioinformatics (Bioconductor)

Mainstream

- ETL
- GIS
- Web
- Image manipulation
- Grab bag (Hex stickers, Image macros)

R Communities

RStudio and Tidyverse

Bioconductor

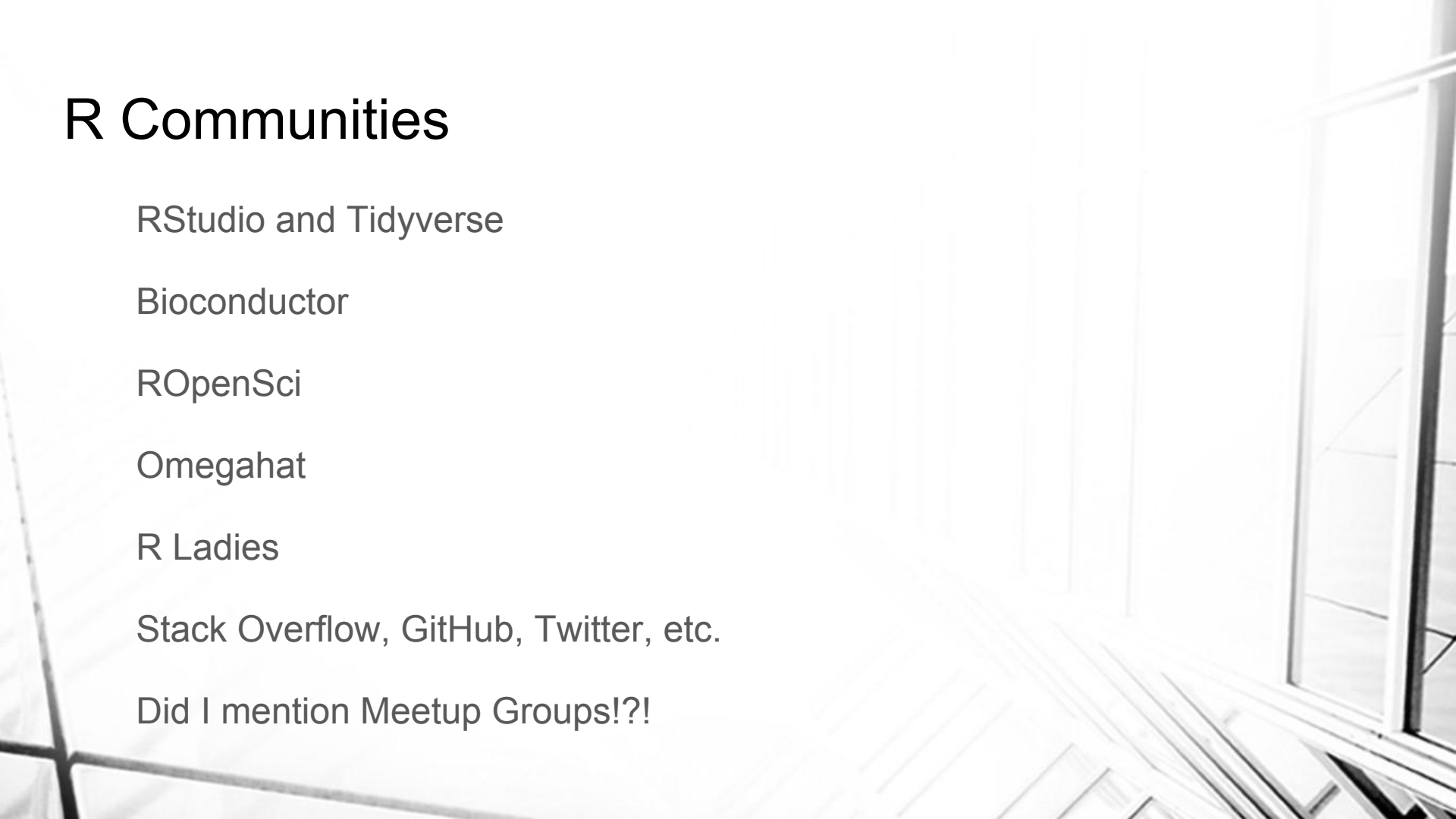
ROpenSci

Omegahat

R Ladies

Stack Overflow, GitHub, Twitter, etc.

Did I mention Meetup Groups!?!



Demo

BIG DATA IGNITE

2nd Lightning Talk

R for Report Automation

-Aaron Clark, Advanced Radiology Services

A Typical “Old School” Reporting Process

Download/ Query Data. Plop into excel for cleaning, summarizing, & graphics

Copy and paste tables and graphics into Word. Perhaps add some notes/narration.

Print as PDF!

Compose and email and attach PDF.

Script the Process in R

Why Change?

- Weave together a narrative of text, graphics, and code, so everything's in one place file!
- Makes analysis very shareable/ reproducible
- You can Scale the process up! More reports != more work
- Use multiple languages, like Python & SQL
- Supports many output formats: HTML, Word, PDF, Slides, etc
- Fast
- Easy

A ~~Real World~~ Example

BIG DATA IGNITE

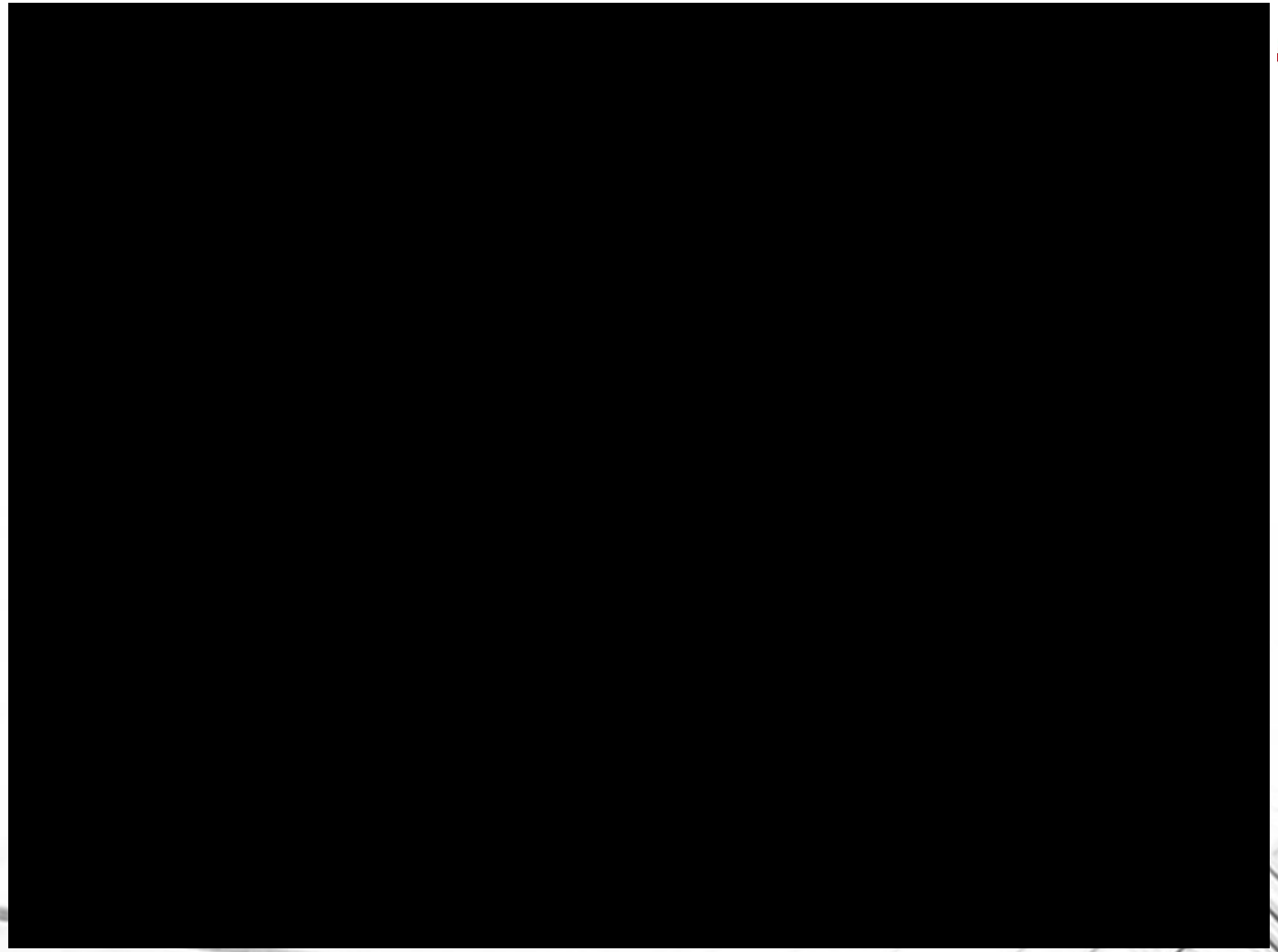
x160+

x32

Demo!

Code: bit.ly/BDI_2018

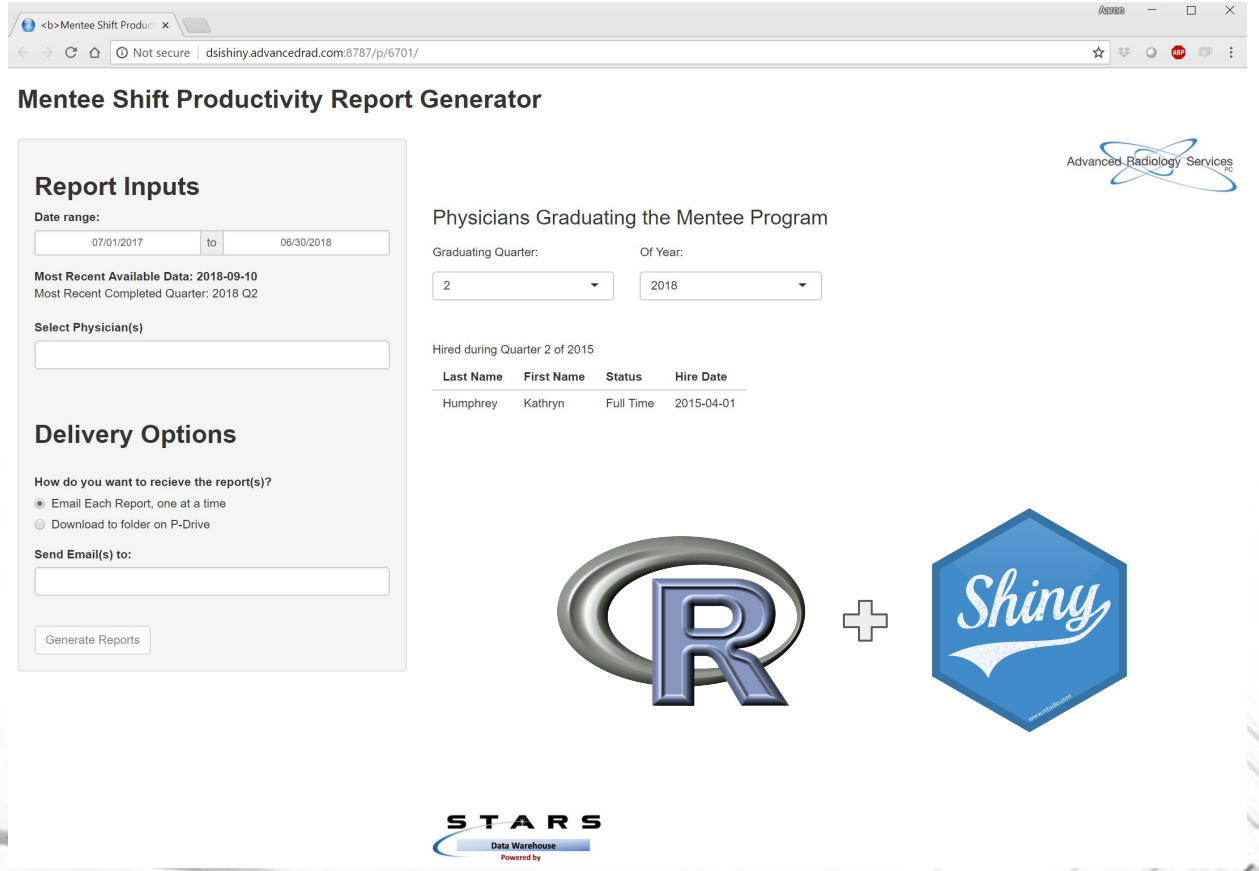
Demo:



BIG DATA IGNITE

With Great Power Comes Great Responsibility

-Ben Parker



Shiny App Report Generator

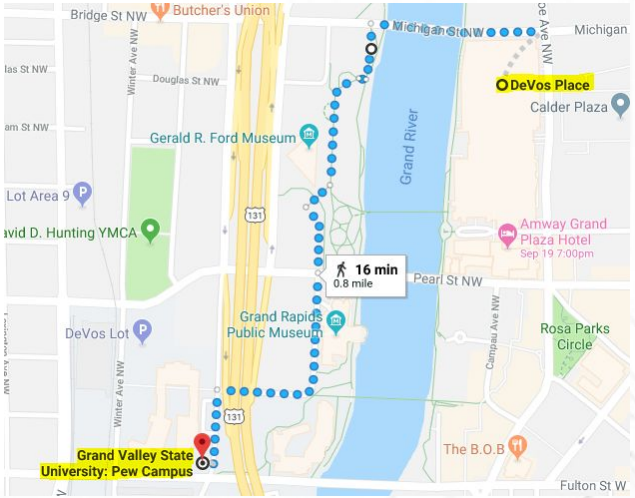
BIG DATA IGNITE

BIG DATA IGNITE

Questions?

Next Meeting: Tues, Oct 16th

R for the Full Analytics Stack

Paul Egeler, Spectrum Health

6p - 8p
DEV 121E
Pew Campus
GVSU