Web Analytics with Hands-on Projects in R

William Shin

Packt Academy

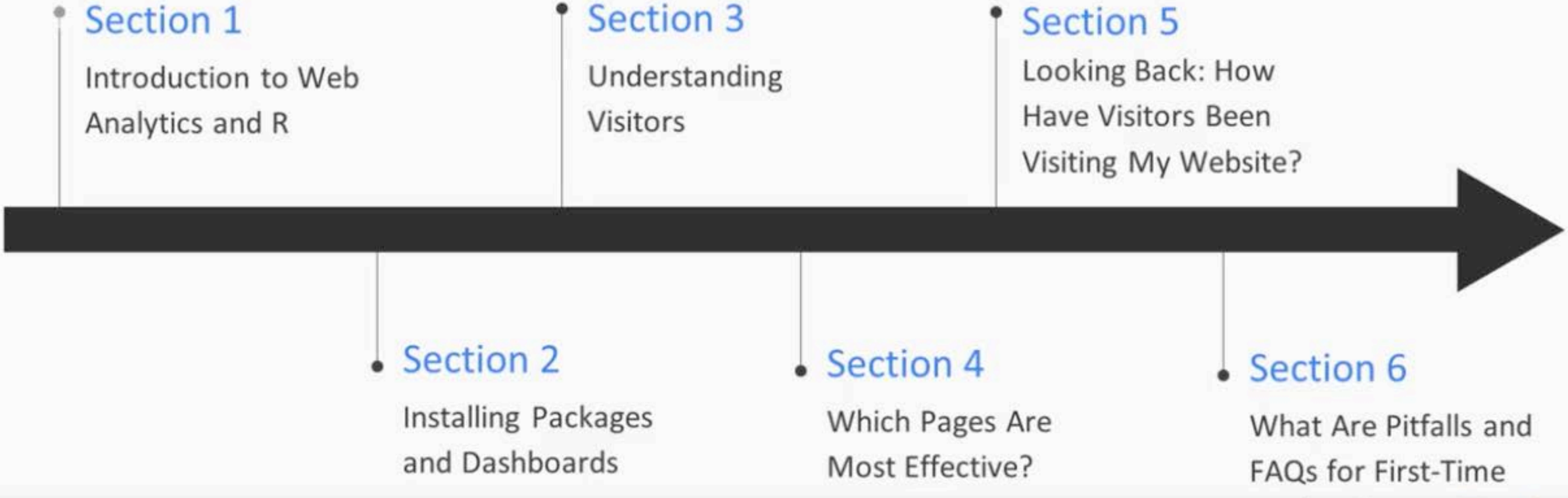
William Shin

- Over 7+ years of experience with R, with extensive work in Google Analytics, R, and plotting over these years
- Experienced leader in analytics and a frequent speaker at scientific conferences and events
- Delivered numerous presentations and analytics solutions to scientific and non-scientific audiences

The Course Overview

Resources available

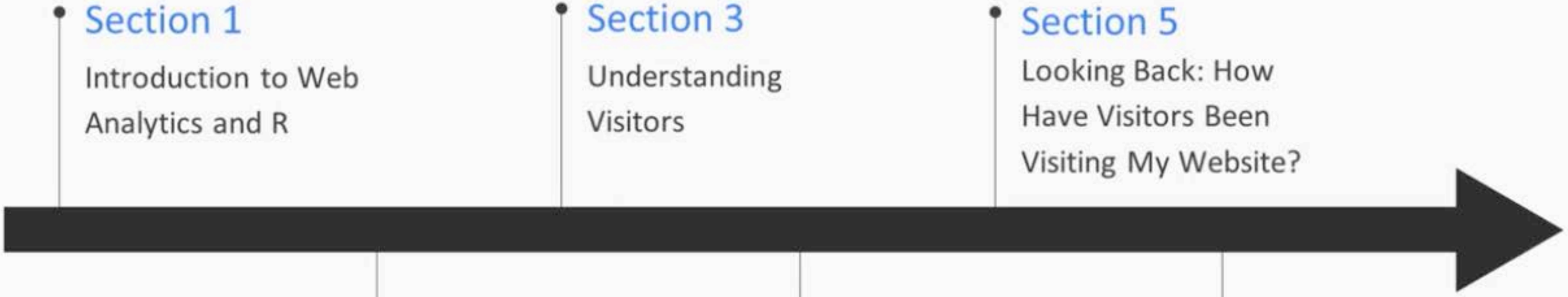
What We'll Learn?



Key Performance Indicators (KPIs) and metrics

- Page Views
- Path Analysis
- Daily Unique Visitors
- Top Exit pages
- Bounce Rates
- Visitor Locations
- Etc.

What We'll Learn?



CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Friday 2017-06-30, Single Candle) R-3.4.1.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.

Submitting to CRAN

To "submit" a package to CRAN, check that your submission meets the CRAN Repository Policy and then use the web form.

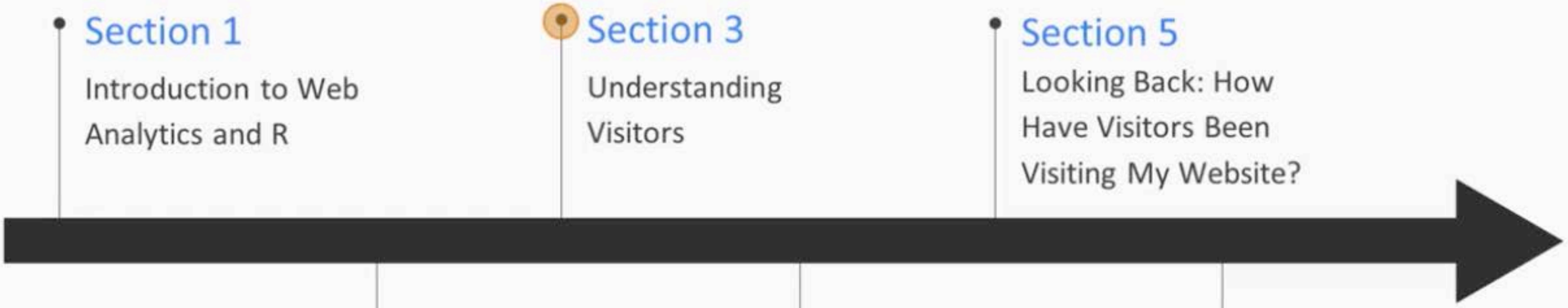
If this fails, upload to ftp://CRAN.R-project.org/incoming/ and send an email to CRAN@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

```
# install.packages("googleAuthR")
# install.packages("googleAnalyticsR")
# install.packages("ggplot2")

library("googleAuthR")
library("googleAnalyticsR")
library("ggplot2")
## authenticate, or use the RStudio Addin "Google API Auth" with analytics scopes
set
ga_auth()

## get your accounts
account_list <- ga_account_list()
account_list$viewId
ga_id <- 154893193

temp_ga_data = google_analytics_4(ga_id,
                  date_range = c("2017-02-22", "2017-05-22"),
                  metrics = c("sessions"),
                  dimensions = c("date"))
```

```
> library("ggplot2")
> ## authenticate, or use the RStudio Addin "Google API Auth" with analytics scopes set
> ga_auth()
Waiting for authentication in browser...
Press Esc/Ctrl + C to abort

>
```

Name	Size	Modified
..		
2.1_WS_Draft.pptx	1.1 MB	Jul 11, 2017, 4:57 PM
2.2_WS_Draft.pptx	109.6 KB	Jul 11, 2017, 9:25 PM
2.3_WS_Draft.pptx	67.4 KB	Jul 11, 2017, 9:48 PM
2.4_WS_Draft.pptx	74.3 KB	Jul 11, 2017, 9:53 PM
2_WS_Draft_dashboard.R	660 B	May 25, 2017, 11:19 PM
2_WS_Draft_googleAnalyticsR.R	1.1 KB	Jul 28, 2017, 11:04 AM
Section2_Metadata.docx	23.9 KB	Jul 18, 2017, 5:01 PM

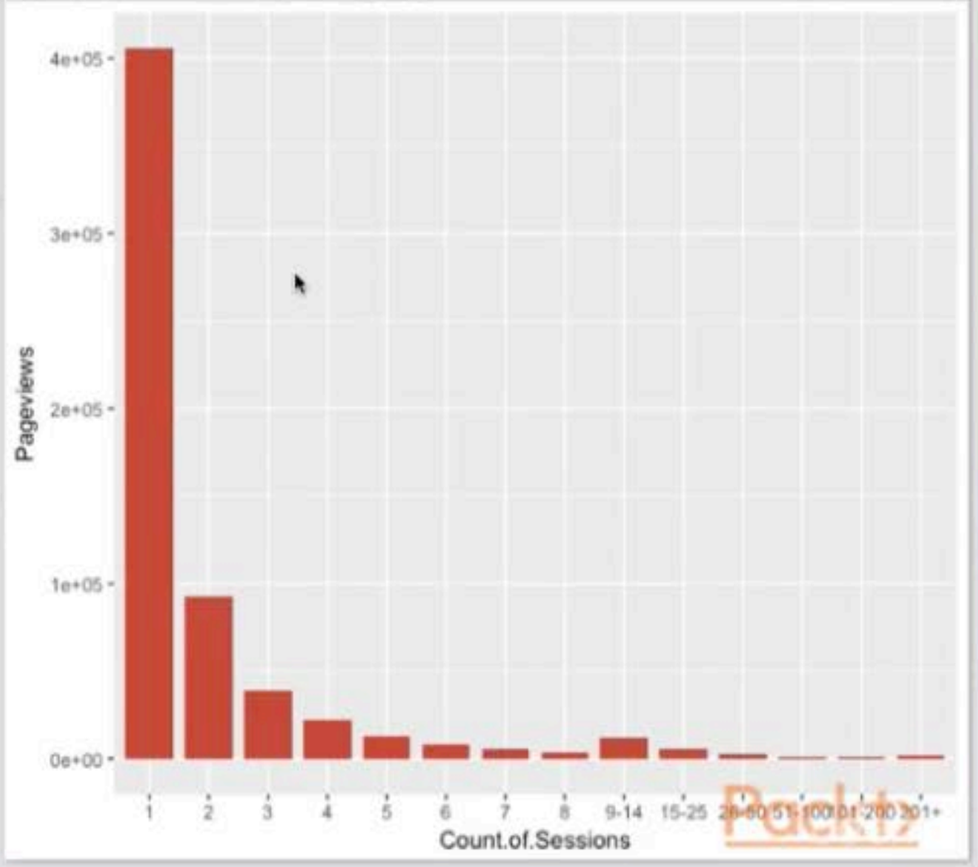
What We'll Learn?



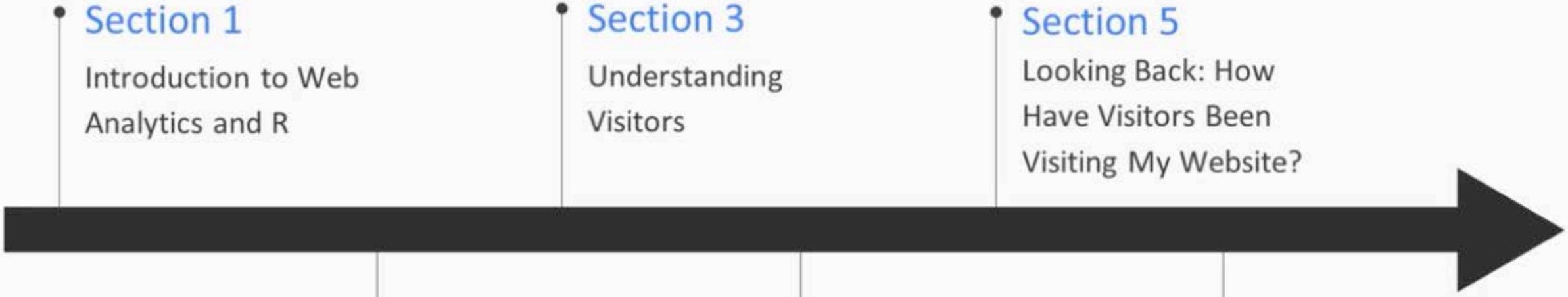
```
ga_data = google_analytics_4(ga_id, date_range = c("2017-04-18", "2017-06-18"),
metrics=c("ga:pageviews", "ga:users"), dimensions=c("ga:sessionCount"))



# using example file
ga_data = read.csv("ExampleFile_3.3_FrequencyRecency.csv", sep=",", header=T)



ga_data.toPlot = data.frame(ga_data)
ga_data.toPlot$Pageviews = as.numeric(gsub(",","", ga_data.toPlot$Pageviews))

#ordering data
ga_data.toPlot$Count.of.Sessions =  factor(ga_data.toPlot$Count.of.Sessions, levels
= ga_data.toPlot$Count.of.Sessions)

# Draw plot
ggplot(ga_data.toPlot, aes(x=Count.of.Sessions, y=Pageviews)) + geom_bar(stat
="identity", width=0.8, fill="tomato3")


```

```
> ga_data.toPlot$Count.of.Sessions =  factor(ga_data.toPlot$Count.of.Sessions, levels = g
a_data.toPlot$Count.of.Sessions)
>
> # Draw plot
> ggplot(ga_data.toPlot, aes(x=Count.of.Sessions, y=Pageviews)) + geom_bar(stat="identity
", width=0.8, fill="tomato3")
>
```

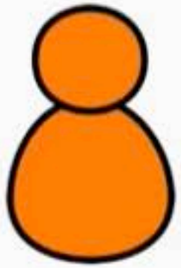
Data	
ga_data	14 obs. of 3 variables
ga_data.toPlot	14 obs. of 3 variables



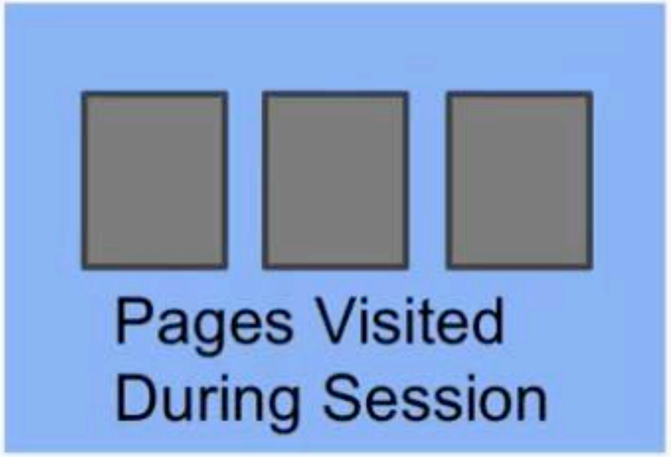
What We'll Learn?



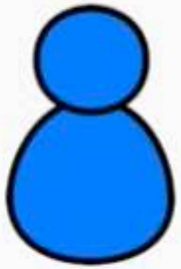
Exit Rates : Measure the percentage that a page is the last page in a session



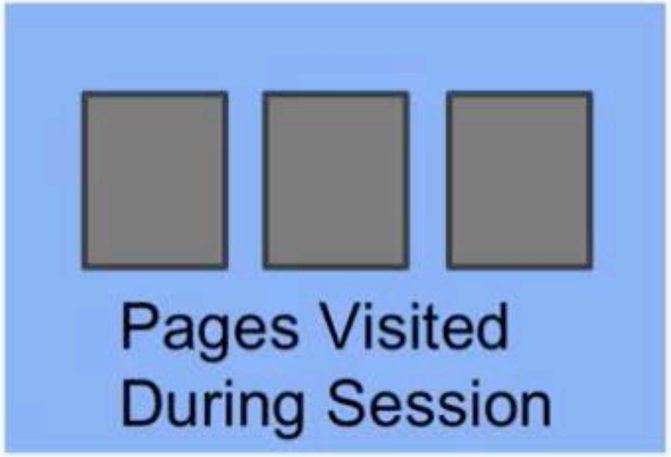
User 1



Bounce Rates : Measure the rate that a page is the only page viewed during a session

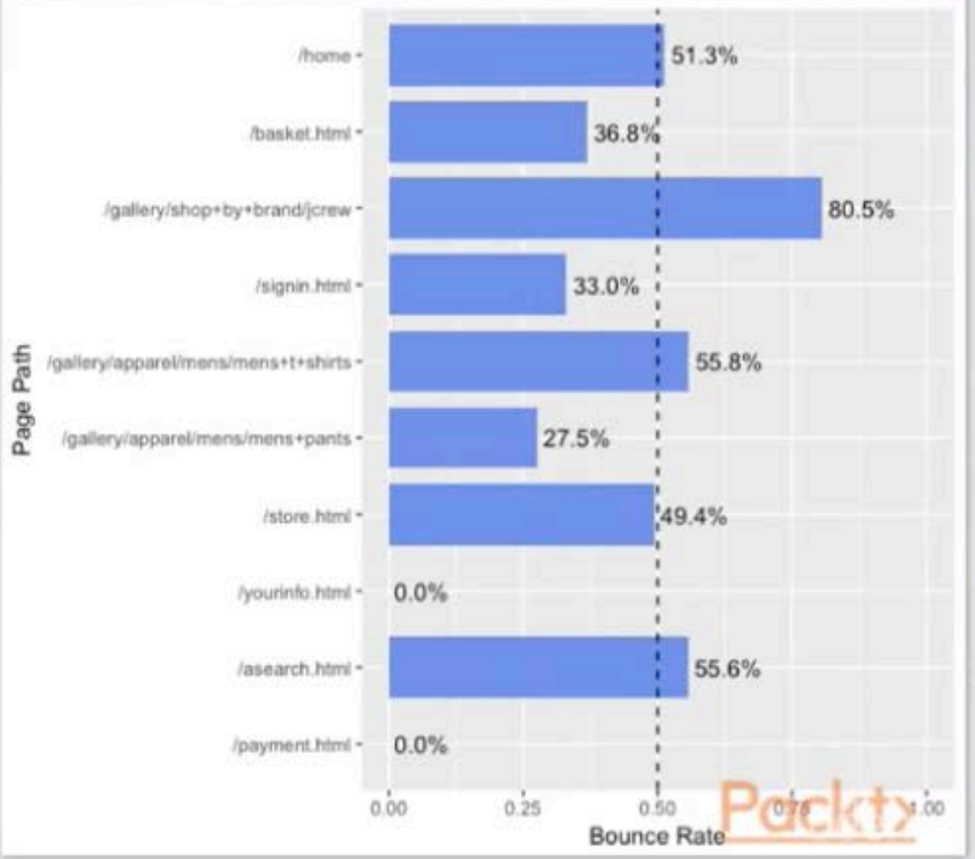


User 2

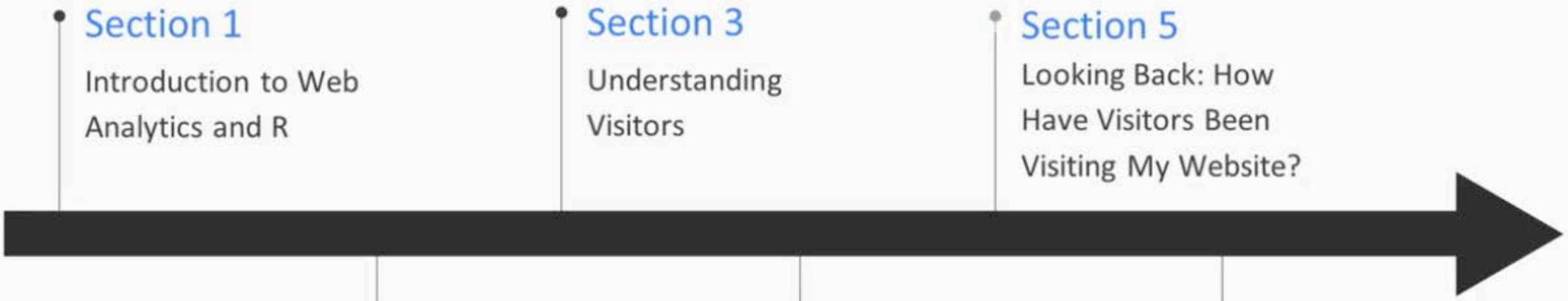


```
# theme_set(theme_bw())
# my.plot = ggplot(ga_data.toPlot, aes(x=Page, y=Bounce.Rate, fill= Bounce.Rate < 0.5)) + geom_bar(stat="identity", width=.8, show.legend = FALSE) + coord_flip()  + ylim(0, 1.0) +geom_text(aes(y = Bounce.Rate, label = percent(Bounce.Rate)),  hjust = -0.1) +
#   geom_hline(yintercept = 0.5,linetype="dashed" ) + ylab("Bounce Rate") + xlab("Page Path")


bounce.Rate.Plot = ggplot(ga_data.toPlot, aes(x=Page, y=Bounce.Rate)) + geom_bar(fill="cornflowerblue",stat="identity", width=.8, show.legend = FALSE) + coord_flip()  + ylim(0, 1.0) +geom_text(aes(y = Bounce.Rate, label = percent(Bounce.Rate)), hjust = -0.1) +
  geom_hline(yintercept = 0.5,linetype="dashed" ) + ylab("Bounce Rate") + xlab("Page Path")
bounce.Rate.Plot
```

```
> bounce.Rate.Plot = ggplot(ga_data.toPlot, aes(x=Page, y=Bounce.Rate)) + geom_bar(fill="cornflowerblue",stat="identity", width=.8, show.legend = FALSE) + coord_flip()  + ylim(0, 1.0) +geom_text(aes(y = Bounce.Rate, label = percent(Bounce.Rate)),  hjust = -0.1) +
+   geom_hline(yintercept = 0.5,linetype="dashed" ) + ylab("Bounce Rate") + xlab("Page Path")
> bounce.Rate.Plot
>
```



What We'll Learn?



Section 1

Introduction to Web Analytics and R

Section 2

Installing Packages and Dashboards

Section 3

Understanding Visitors

Section 4

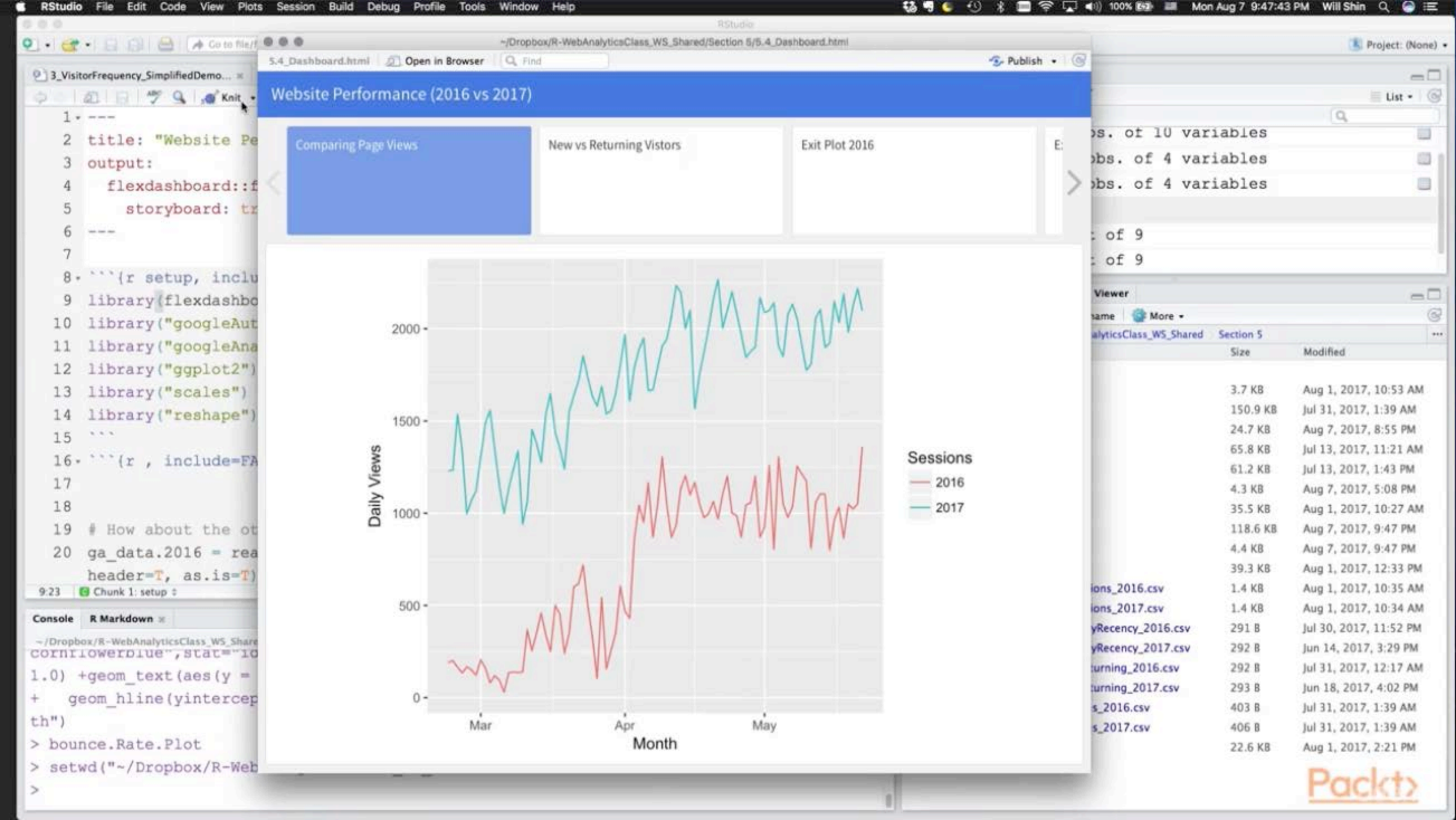
Which Pages Are Most Effective?

Section 5

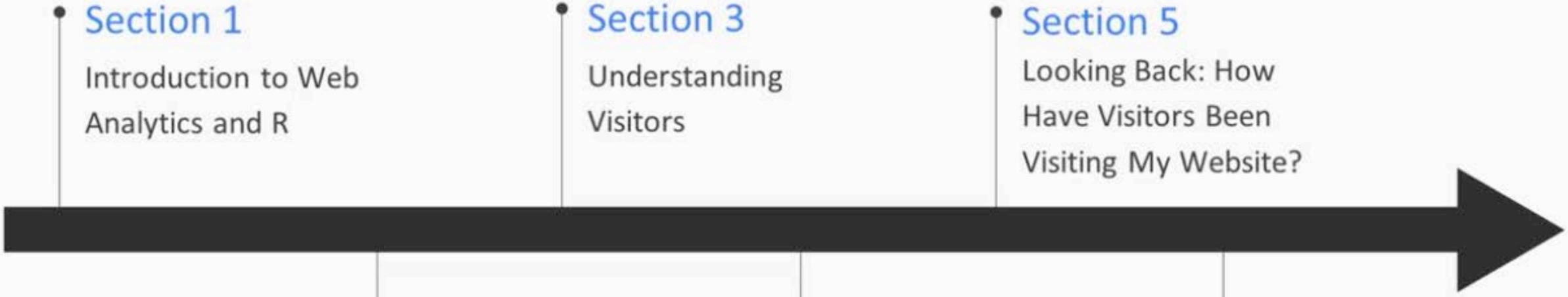
Looking Back: How Have Visitors Been Visiting My Website?

Section 6

What Are Pitfalls and FAQs for First-Time Analysts?


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
Mon Aug 7 9:47:43 PM Will Shin
~/Dropbox/R-WebAnalyticsClass_WS_Shared/Section 5/5.4_Dashboard.html
5.4_Dashboard.html
Open in Browser
Find
Publish
Website Performance (2016 vs 2017)
Comparing Page Views
New vs Returning Vistors
Exit Plot 2016
Daily Views
2000
1500
1000
500
0
Mar
Apr
May
Month
Sessions
2016
2017
3_VisitorFrequency_SimplifiedDemo...
Knit
1 ---
2 title: "Website Pe
3 output:
4 flexdashboard::f
5 storyboard: tr
6 ---
7
8 ```{r setup, inclu
9 library(flexdashbo
10 library("googleAut
11 library("googleAna
12 library("ggplot2")
13 library("scales")
14 library("reshape")
15 ```
16 ```{r , include=FA
17
18
19 # How about the ot
20 ga_data.2016 = rea
header=T, as.is=T)
9:23
Chunk 1: setup
Console
R Markdown
~/Dropbox/R-WebAnalyticsClass_WS_Share
1.0) +geom_text(aes(y =
+ geom_hline(yintercep
th")
> bounce.Rate.Plot
> setwd("~/Dropbox/R-Web
>
Project: (None)
List
obs. of 10 variables
obs. of 4 variables
obs. of 4 variables
of 9
of 9
Viewer
More
alyticsClass_WS_Shared
Section 5
Size
Modified
3.7 KB
Aug 1, 2017, 10:53 AM
150.9 KB
Jul 31, 2017, 1:39 AM
24.7 KB
Aug 7, 2017, 8:55 PM
65.8 KB
Jul 13, 2017, 11:21 AM
61.2 KB
Jul 13, 2017, 1:43 PM
4.3 KB
Aug 7, 2017, 5:08 PM
35.5 KB
Aug 1, 2017, 10:27 AM
118.6 KB
Aug 7, 2017, 9:47 PM
4.4 KB
Aug 7, 2017, 9:47 PM
39.3 KB
Aug 1, 2017, 12:33 PM
ions_2016.csv
1.4 KB
Aug 1, 2017, 10:35 AM
ions_2017.csv
1.4 KB
Aug 1, 2017, 10:34 AM
yRecency_2016.csv
291 B
Jul 30, 2017, 11:52 PM
yRecency_2017.csv
292 B
Jun 14, 2017, 3:29 PM
urning_2016.csv
292 B
Jul 31, 2017, 12:17 AM
urning_2017.csv
293 B
Jun 18, 2017, 4:02 PM
s_2016.csv
403 B
Jul 31, 2017, 1:39 AM
s_2017.csv
406 B
Jul 31, 2017, 1:39 AM
22.6 KB
Aug 1, 2017, 2:21 PM
Packt>

What We'll Learn?



Ways to prevent analysis paralysis

- Differentiate between “small decisions” and “large decisions”
- Focus on the results and metrics that matter the most.
- Focus on small actions first.
- Simplify, simplify, simplify!

Course Goals

- Track, measure, and analyze your website performance effectively using popular tools in R
- Get the most out of R's analytics capabilities and generate insightful reports and visualization
- A comprehensive tutorial to performing efficient web analytics with R to increase your website’s effectiveness

Prerequisites

- Some programming experience (in any language) is recommended, but not required
- A basic understanding of web analytics concepts will be helpful, but not required