

Introduction to Data Analysis & Visualization with R

Brian Zelip

Emerging Technologies Librarian

bzelp@hshsl.umaryland.edu

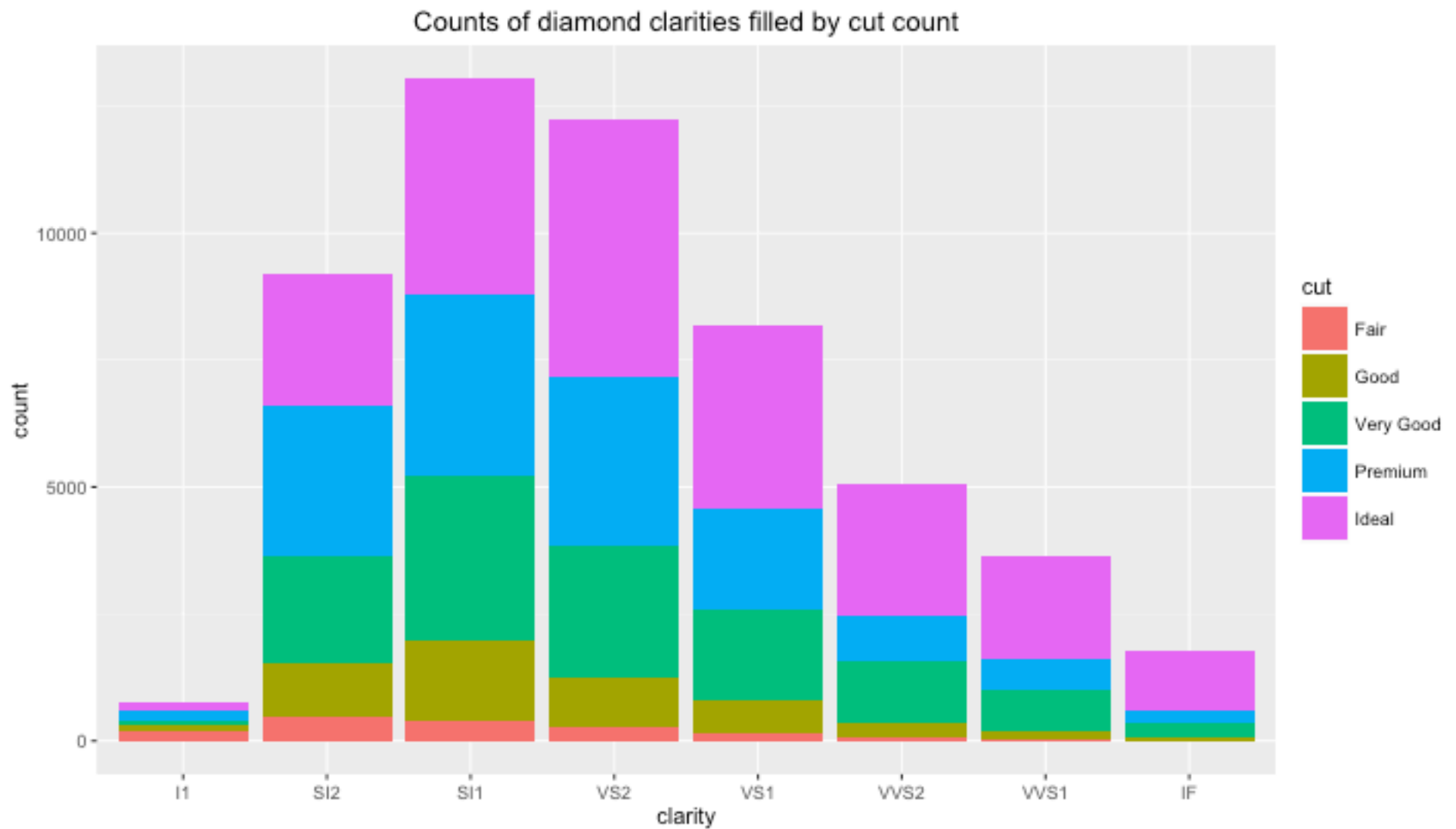
Agenda of Workshop

1. R basics
2. Installation
3. Practice
 1. Plotting data
 2. Statistical analysis

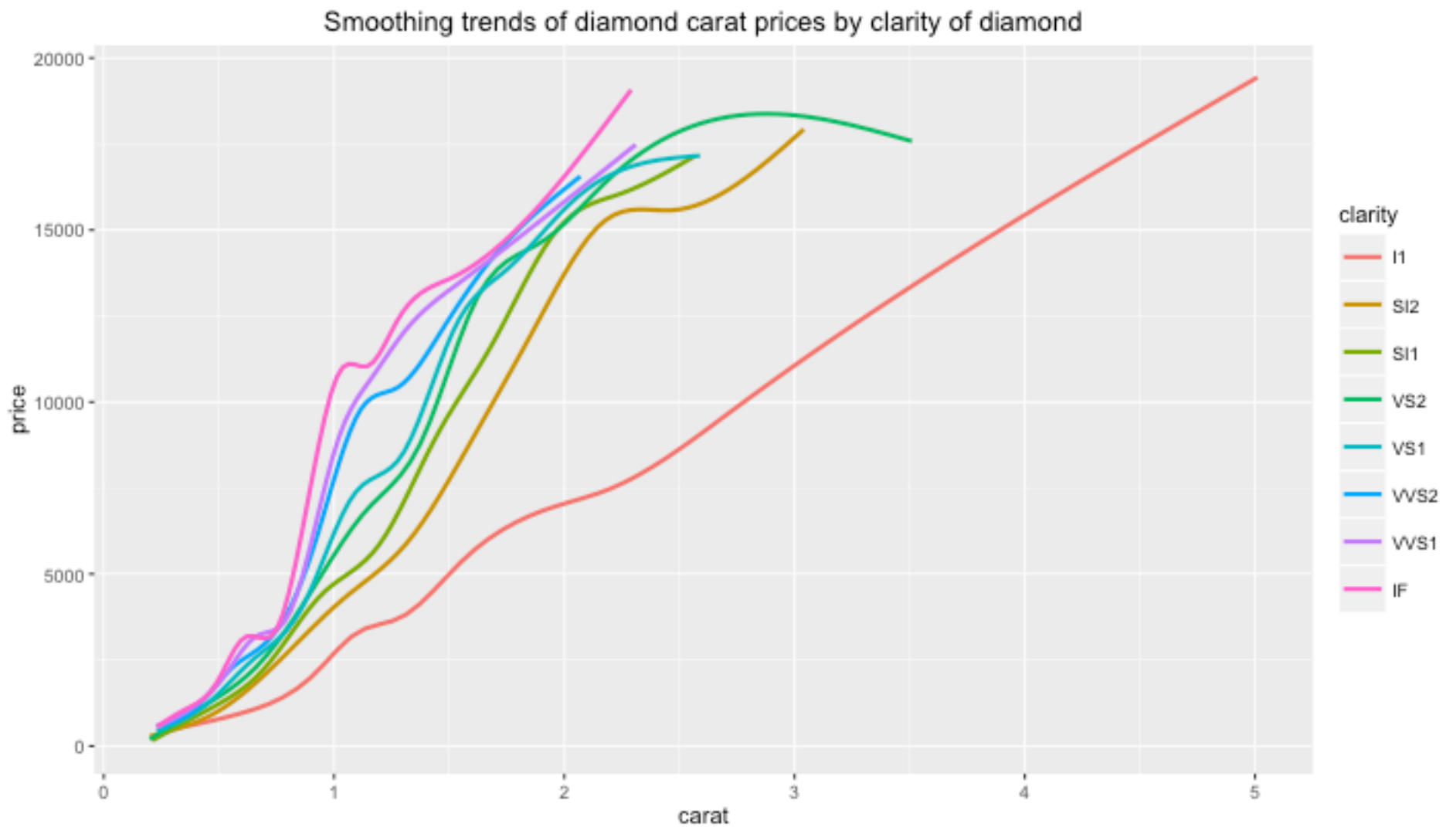
Scope of Workshop

- Demonstration over explanation of statistical concepts

Examples of Workshop Output



Examples of Workshop Output



Examples of Workshop Output

```
> fit = lm(mpg ~ wt, mtcars)
> summary(fit)
```

Call:

```
lm(formula = mpg ~ wt, data = mtcars)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.5432	-2.3647	-0.1252	1.4096	6.8727

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.2851	1.8776	19.858	< 2e-16 ***
wt	-5.3445	0.5591	-9.559	1.29e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.046 on 30 degrees of freedom

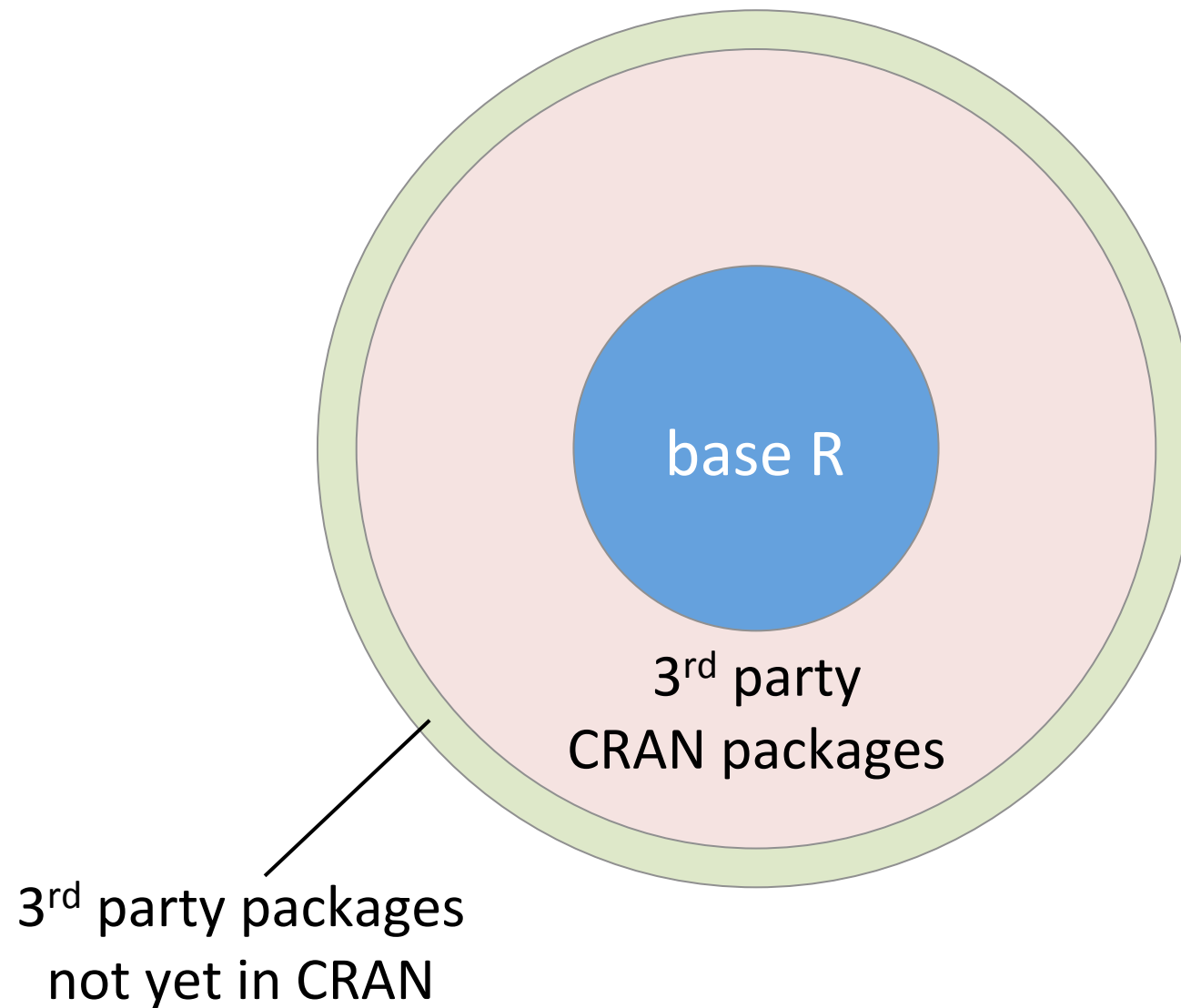
Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446

F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

About R

- <https://www.r-project.org/>
- Free and open source software
- Runs on Windows, OSX, GNU/Linux, FreeBSD, and more
- R is a programming language and environment
- Default install provides “base R” suite of packages, or tools
- Install more packages via CRAN
 - CRAN = Comprehensive R Archive Network
 - <https://cran.r-project.org/>
 - The main source for downloading for R and most packages
 - Each package comes with documentation
 - Very active global community of users and developers (so help and guidance can usually be found quickly online)

R's Package Ecosystem



RStudio

- <https://www.rstudio.com/>
- Free and open source software
- Combines the power of the command line with the usability of a graphical user interface
- Provides the same experience across operating systems (which the R application and other GUIs do not)
- Requires prior R installation

Brief Look at R Syntax for Basic Calculation

```
> #  
> # Basic arithmetic  
> #  
>  
> 3 + 5      # addition  
[1] 8  
> 20 - 5     # subtraction  
[1] 15  
> 10 * 10    # multiplication  
[1] 100  
> 3 / 4      # division  
[1] 0.75  
> 18 %/% 12   # just the integer part of the quotient  
[1] 1  
> 18 %% 12    # just the remainder part (modulo)  
[1] 6  
> 10 ^ 2     # exponentiation  
[1] 100  
> log(10)    # natural log (base e)  
[1] 2.302585  
> exp(2.302585) # antilog, e raised to a power  
[1] 9.999999  
> log10(100) # base 10 logs; log(100, base=10) is the same  
[1] 2  
> sqrt(100)  # square root  
[1] 10  
> |
```

Brief Look at More Useful R Syntax

Working with variables

```
x <- 5
```

```
y <- 9
```

```
y - x
```

```
[1] 4
```

```
sqrt(y)
```

```
[1] 3
```

```
data <- read.csv("myData.csv", header = T)
```

Keyword functions

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

```
# Install the ggplot2 library
```

```
library("ggplot2")
```

```
# Load the ggplot2 library into current workspace
```

```
library()
```

```
# List all the packages built in to base R
```

```
data()
```

```
# List all the datasets built in to base R
```

```
help(mtcars)
```

```
# Show the description of the mtcars dataset
```

```
example("ggplot2")
```

```
# Show examples of using the ggplot2 library
```

```
getwd()
```

```
# Get the current working directory for current workspace
```

```
setwd()
```

```
# Set the current working directory for current workspace
```

```
list.files()
```

```
# List the files in the working directory
```

INSTALLATION

1. R, <https://cran.r-project.org/>
2. RStudio, <https://www.rstudio.com/>

HANDS-ON DEMO



UNIVERSITY of MARYLAND
Health Sciences and
Human Services Library

Library Guide on R

<http://guides.hshsl.umaryland.edu/R>

Search the Library Website

Resources ▾ Services ▾ Assistance ▾ About the Library ▾

OneSearch Ask Us! Hours Follow Us ▾

[HS/HSL](#) / [Guides](#) / [R](#) / Recommended R Resources

R

An introduction to the R programming language.

Introduction to R

Installing R and R Packages

Recommended R Resources

RECOMMENDED R RESOURCES

Up and running with R

- [R website](#)
- [CRAN, Comprehensive R Archive Network -download R and additional 3rd party packages](#)
- [RStudio - a 3rd party integrated development environment for R \(install R before RStudio\)](#)
- [All R packages available from CRAN, sorted by name](#)
- [All R packages available from CRAN, sorted by date of publication](#)
- [Browse CRAN packages by topic and tools to automatically install all packages for special areas of interest](#)

Packages for common needs

Data wrangling