Introduction to R
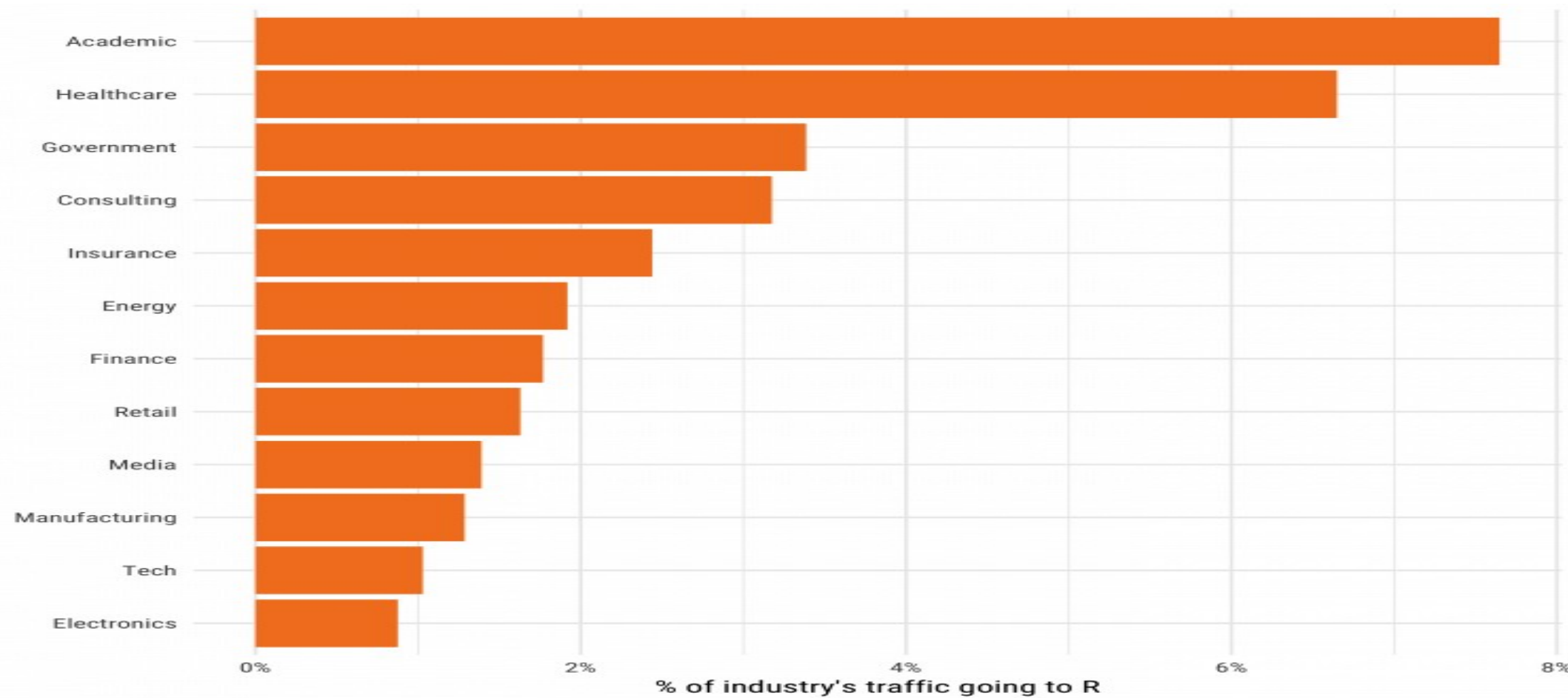
➔ What is R ?

➔ R is a programming language and free software environment

➔ Primarily used for statistical computing and graphics.

➔ It provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, time-series analysis, classification, clustering, and more.

Who is use R ?

% of industry's traffic going to R

Why learning R?

➜ R is open source, so it's free.
➜ R is cross-plateform compatible, so it can be installed on Windows, MAC OSX an
  Linux
➜ R provides a wide variety of statistical techniques and graphical capabilities.
➜ R provides the possibility to make a reproducible research by embedding script an
  results in a single file.
➜ R is highly extensible and it has thousands of well-documented extensions (named
  packages) for a very broad range of applications in the statistical analysis, healt
  care,…

Rstudio

➔ R Studio is an integrated development environment (IDE) that supports the R programming language.

➔ It has a user-friendly interface for writing, running, debugging, and visualizing R code

Source Pane
Edit and run scripts (e.g. Rmarkdown templates), and view datasets

Environment Pane
Overview of objects (datasets, parameters, lists, etc.) you have imported or created.

Tip: Run script

Tip:
Start new
script

Tip: Zoom
and export
plots

R Console Pane
R commands run are shown here, and non-graphic output and errors are displayed

Plots, Packages, and Help Pane
Commonly used to view graphics, install packages, and view help

C:/Users/Neale/OneDrive - Neale Batra/Desktop/R4EPIs-website - master - RStudio

File   Edit   Code   View   Plots   Session   Build   Debug   Profile   Tools   Help

Go to file/function   Addins

R4EPIs-website — Desktop

RStudioOverview.Rmd   skeleton.Rmd   linelist_cleaned

Knit   Insert   Run

```
1  ---
2  title: "Cholera outbreak report"
3  output:
4    word_document:
5      keep_md: true
6  ---
7
8  # Introduction to this template
9
10 This is a template which can be used to create an automated outbreak situation
11 report for cholera.
12
13 - It is organised by time, place and person.
14 - You can type normal text in white spaces (such as here) and r-code in grey
15   spaces (denoted by three backticks and r) (see [Rmarkdown
16   introduction](https://rmarkdown.rstudio.com/articles_intro.html) and
17   [Markdown basics](https://rmarkdown.rstudio.com/authoring_basics.html))
18 - Introductions and contents of sections are within square brackets "[...]" and
19   can be deleted as appropriate
```

767:1   Demographics   R Markdown

Environment   History   Connections   Build   Git

Import Dataset   List

Global Environment

linelist_clea...   300 obs. of 54 variables
linelist_dict   183 obs. of 11 variables
linelist_raw   300 obs. of 46 variables
population_da...   5 obs. of 3 variables
population_da...   4 obs. of 3 variables
population_da...   4 obs. of 3 variables

Values
first_week   "2018-w01"
LABS   chr [1:3] "cholera_culture_result" "ch...
obs_end   2018-05-20
obs_start   2018-01-01

Files   Plots   Packages   Help   Viewer

Zoom   Export

Console   Terminal

C:/Users/Neale/OneDrive - Neale Batra/Desktop/R4EPIs-website/

```
> plot_age_pyramid(linelist_cleaned,
+                  age_group = "age_group",
+                  split_by = "sex",
+                  stack_by = "case_def") +
+   labs(y = "Cases (n)", x = "Age group (years)") + # change axis labels (nb. x/y flip)
+   theme(legend.position = "bottom",      # move legend to bottom
+         legend.title = element_blank(), # remove title
+         text = element_text(size = 18)   # change text size
+     )
+ ...
>
Error: attempt to use zero-length variable name
>
```

Male   Female   nown/unspec

Age group (years)

45+
30-44
15-29
5-14
0-4

Cases (n)

Confirmed   Probable   Suspected

R Packages

➔ R packages consist of R functions, compiled code, and sample data. In the R environment, they are placed in a directory named "library"

➔ R automatically installs a set of packages during installation

➔ More packages are can be added later, as they are required for a specific purpose

➔ R packages could be easily installed using the default function

install.packages("package_name")

Most used packages in R and why ?

➔ **Ggplot2**: A powerful package for creating high-quality and customizable graphs, It provides a high-level interface for creating a wide range of plots.

➔ **Dplyr**: dplyr: Another package by Hadley Wickham, dplyr provides a set of functions for data manipulation and transformation. It offers intuitive verbs like **filter**(), **select**(), **mutate**(), **group_by**(), and **summarize**() for efficiently handling data manipulation tasks

➔ **Tidyverse**: A collection of packages designed to work together seamlessly for data manipulation, visualization, and analysis in R

R and R-studio installation

R installation instruction

➔ Windows : https://cran.mirror.ac.za/bin/windows/

➔ Linux : https://cran.mirror.ac.za/bin/linux/

➔ McOS : https://cran.mirror.ac.za/bin/macosx/

R studio installation

➔ https://posit.co/download/rstudio-desktop/

Descriptive Statistics and Graphics

http://www.sthda.com/english/wiki/descriptive-statistics-and-graphics

Data visualisation with ggplot2

https://rpubs.com/GeospatialEcologist/DataViz

The Epidemiologist R Handbook

https://epirhandbook.com/en/reports-with-r-markdown.html