



GENERAL ASSEMBLY

Python 101

Joseph Nelson

Founder, Principal Data Scientist, ROC AUC, LLC

ABOUT ME

- **Data Scientist**
- **From:** Des Moines, Iowa
- **Influences:** Jeff Dean, Andrew Ng, Rachel Thomas
- **Likes:** Hockey, SaaS, bad data science puns, cycling



ABOUT YOU

BEFORE WE DIVE IN...

Let's talk a bit about you!

Name

What brings you to GA?

- Current activities
- Future goals

Fun Fact

LEARNING OBJECTIVES

- Discover the history of Python and how it compares to other programming languages.
- Touch on fundamental Python programming techniques and tools.
- Discuss its applications in data science and the types of problems it can solve.
- See where Python programming fits into the data workflow.
- Apply your new skills to solve a real-world problem with Python.

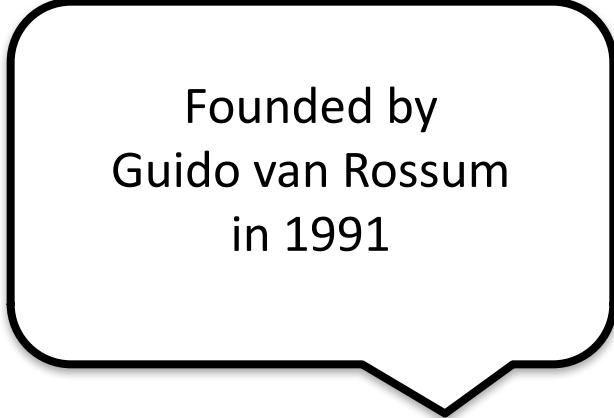
WHAT TOOLS DO DATA SCIENTISTS USE?

**When you hear Python,
what do you think?**

WHAT TOOLS DO DATA SCIENTISTS USE?

WHY PYTHON?

- Created for simplicity and readability
- Rapid prototyping, ease of production
- Open source, importable libraries
- Broad range of applications
- Fast growing community



Founded by
Guido van Rossum
in 1991

WHAT TOOLS DO DATA SCIENTISTS USE?

ZEN OF PYTHON

Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.

Flat is better than nested.
Sparse is better than dense.

Readability counts.

Special cases aren't special enough to break the rules.

Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.

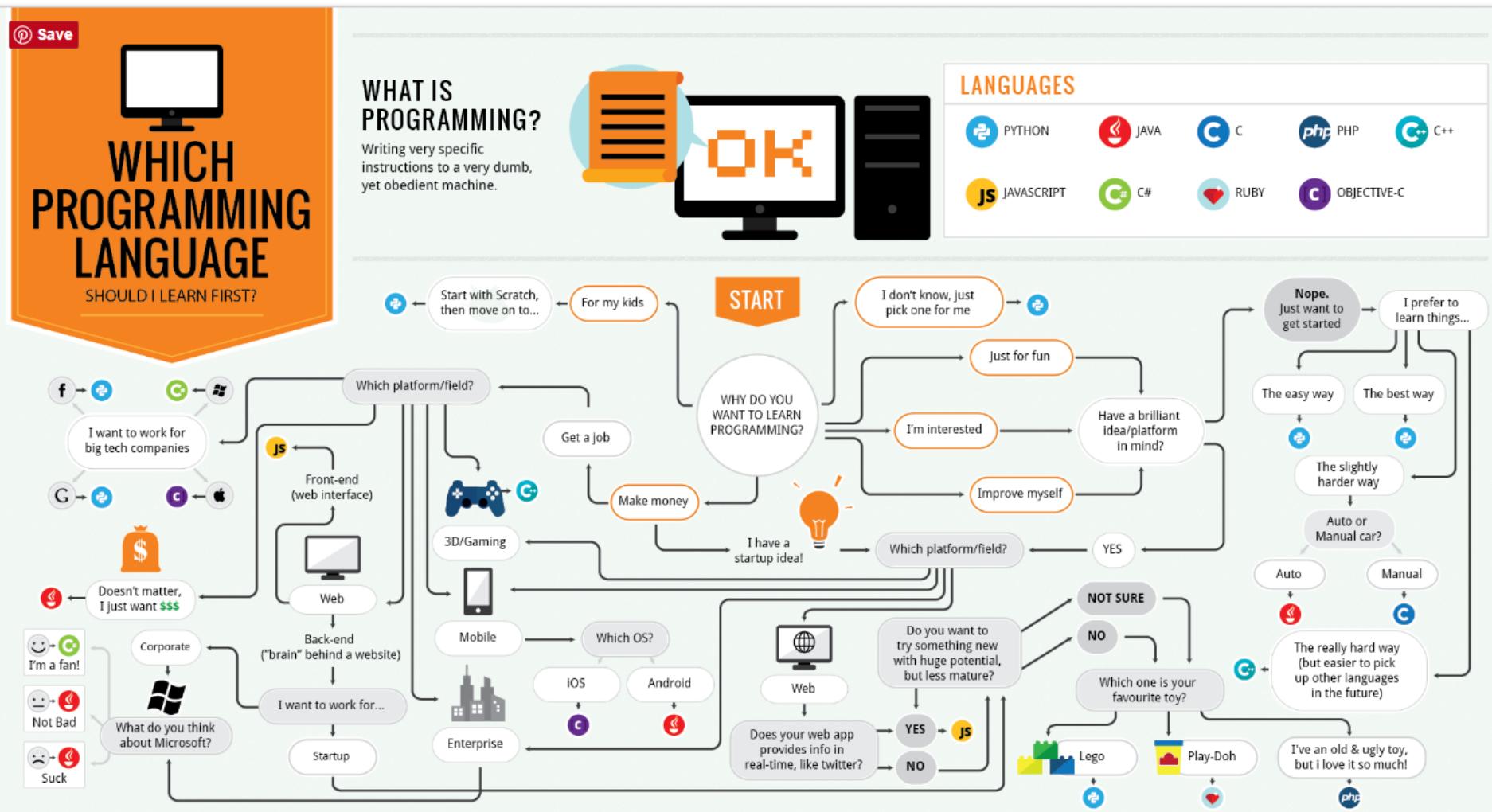
In the face of ambiguity, refuse the temptation to guess.
There should be one—and preferably only one—obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.

Now is better than never.

Although never is often better than **right** now.

If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea—let's do more of those!

WHY PYTHON?



WHY PYTHON?

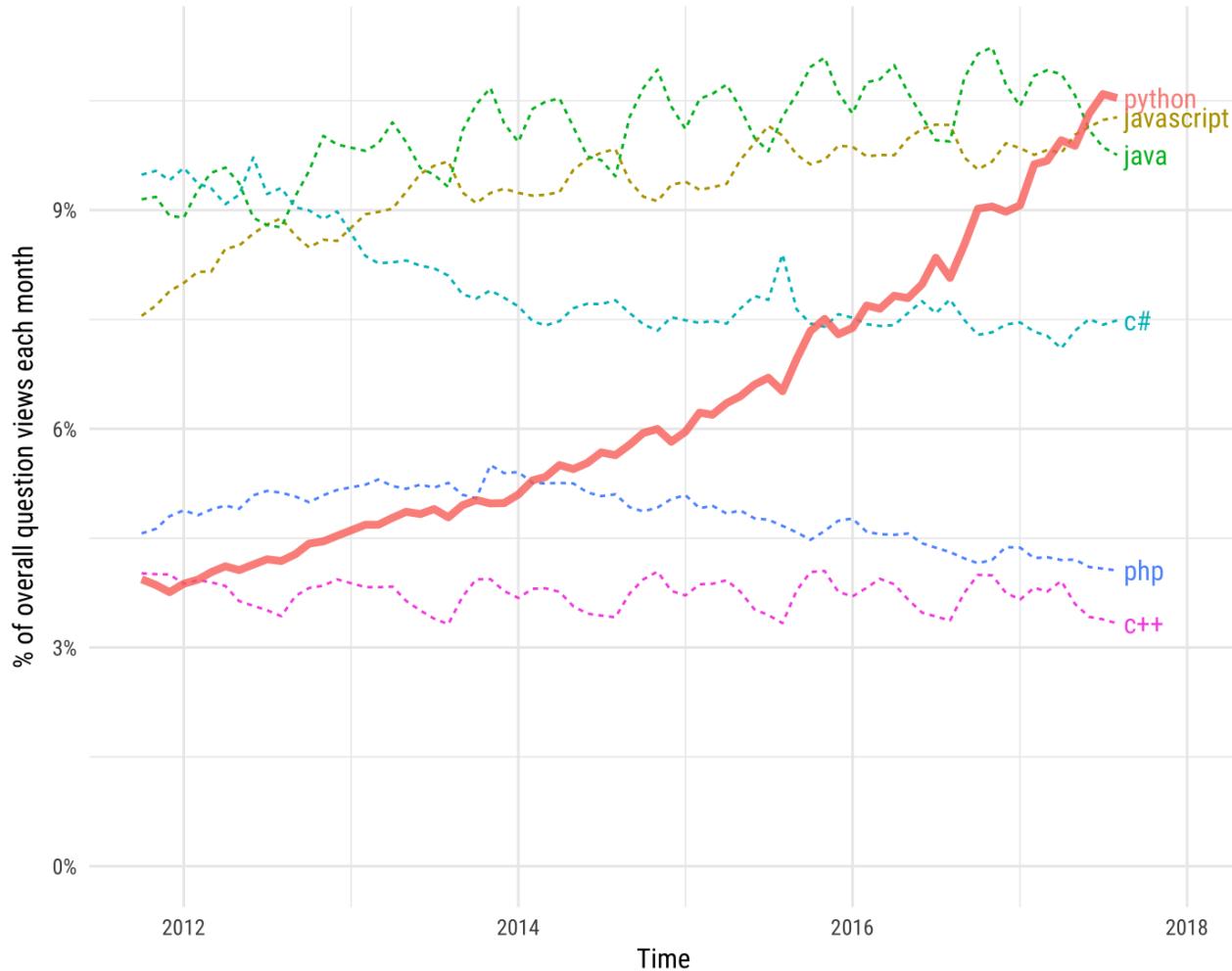


Language Rank	Types	Spectrum Ranking
1. C	📱💻⚙️	100.0
2. Java	🌐📱💻	98.1
3. Python	🌐💻	98.0
4. C++	📱💻⚙️	95.9
5. R	💻	87.9
6. C#	🌐📱💻	86.7
7. PHP	🌐	82.8
8. JavaScript	🌐📱	82.2
9. Ruby	🌐💻	74.5
10. Go	🌐💻	71.9

WHAT TOOLS DO DATA SCIENTISTS USE?

Growth of major programming languages

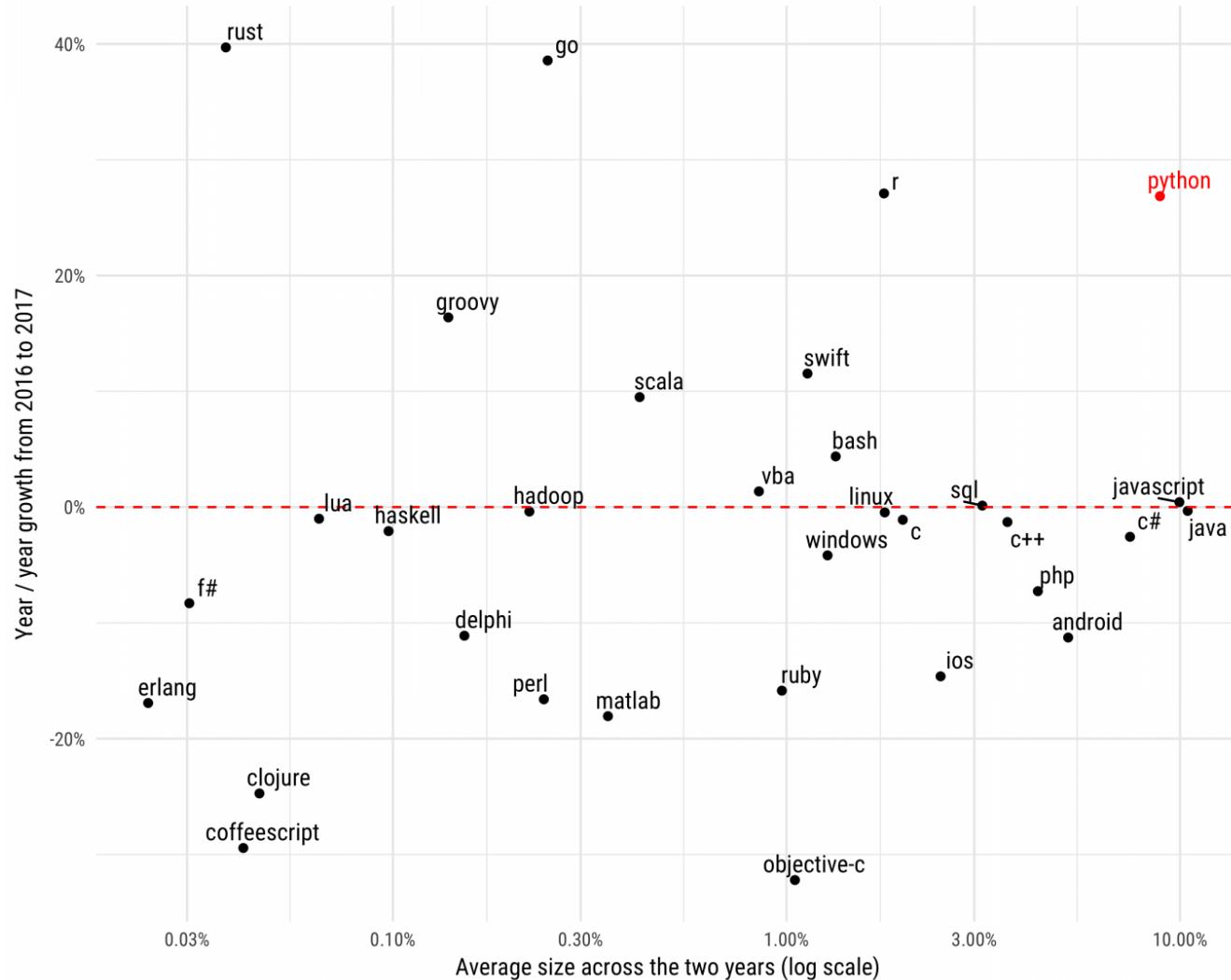
Based on Stack Overflow question views in World Bank high-income countries



WHAT TOOLS DO DATA SCIENTISTS USE?

Year over year growth in traffic to programming languages/platforms

Comparing question views in January-August of 2016 and 2017, in World Bank high-income countries.
TypeScript had a growth rate of 142% and an average size of .36%; and was omitted.



WHY PYTHON?

C

```
#include <cs50.h>

int main(void)
{
    puts("Hello, world!");
}
```

Java

```
import javax.swing.JFrame; //Importing class JFrame
import javax.swing.JLabel; //Importing class JLabel
public class HelloWorld {
    public static void main(String[] args) {
        JFrame frame = new JFrame(); //Creating frame
        frame.setTitle("Hi!"); //Setting title frame
        frame.add(new JLabel("Hello, world!")); //Adding text to frame
        frame.pack(); //Setting size to smallest
        frame.setLocationRelativeTo(null); //Centering frame
        frame.setVisible(true); //Showing frame
    }
}
```

Python

```
print("Hello, world!")
```

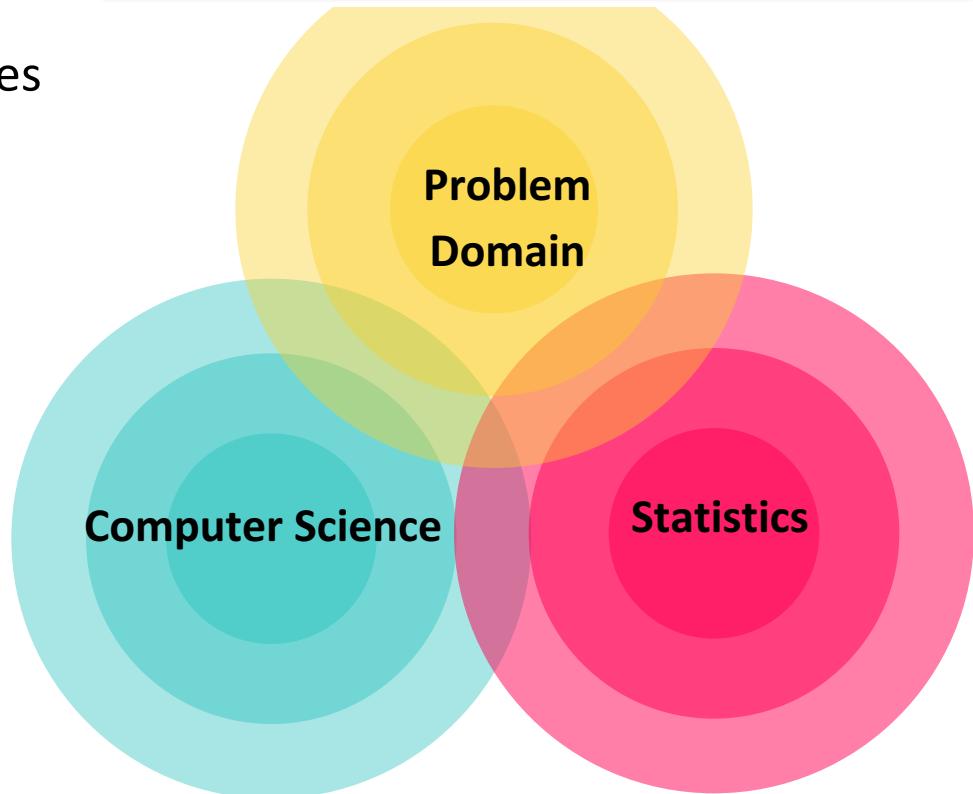
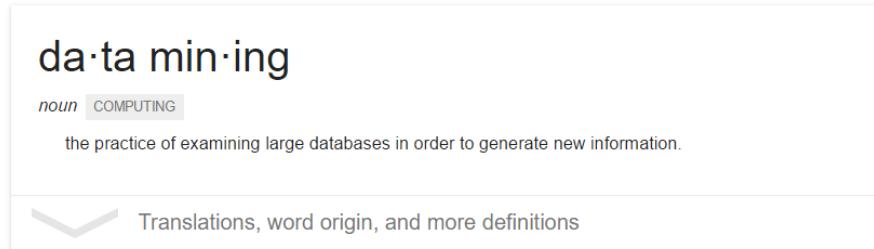
LEARNING OBJECTIVES

- Discover the history of Python and how it compares to other programming languages.
- Touch on fundamental Python programming techniques and tools.
- Discuss its applications in data science and the types of problems it can solve.
- See where Python programming fits into the data workflow.
- Apply your new skills to solve a real-world problem with Python.

WHAT IS A DATA SCIENTIST?

MY DEFINITION

Data Scientists use **data mining** techniques to generate new insights that increase efficiencies / give businesses a competitive advantage



HOW DO DATA SCIENTISTS ADD VALUE?

VALUE ADD

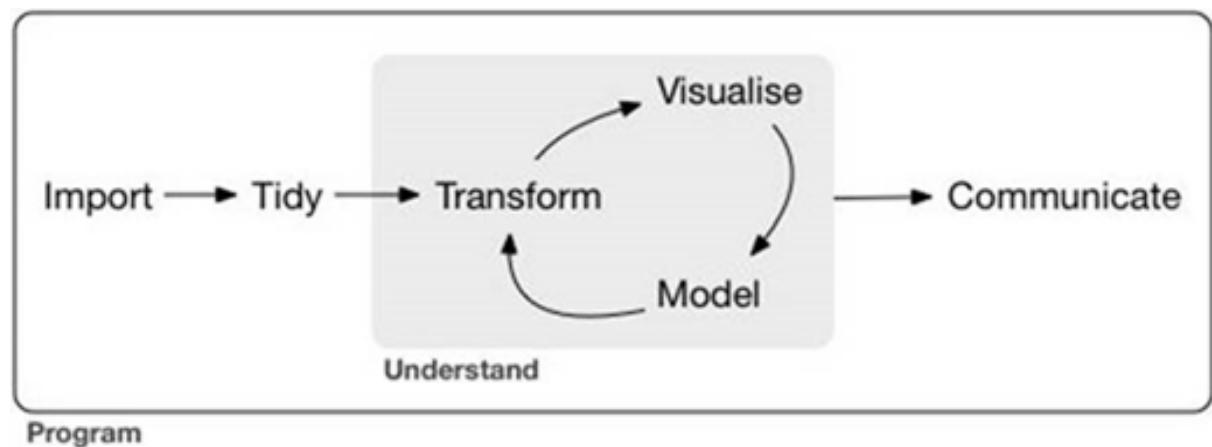
- Predicting the bad
- Identifying the good
- Automating existing processes

“Every good data analysis starts
with a bad data analysis.”
– Hadley Wickham

WHAT IS THE DATA SCIENCE WORKFLOW?

WORKFLOW

- Define the problem / question
- Identify and collect data
- Explore and prepare data
- Build and evaluate model
- Communicate results



WHAT TOOLS DO DATA SCIENTISTS USE?

KEY TOOLS

Querying /
Collection



- Database Queries
 - Web Scraping
 - API Calls

WHAT TOOLS DO DATA SCIENTISTS USE?

KEY TOOLS

Querying /
Collection



- Database Queries
- Web Scraping
 - API Calls

Manipulation /
Modeling



- Munging and wrangling
- Merging and enhancing
 - Building Models

WHAT TOOLS DO DATA SCIENTISTS USE?

KEY TOOLS

Querying /
Collection



- Database Queries
- Web Scraping
 - API Calls

Manipulation /
Modeling



- Munging and wrangling
- Merging and enhancing
 - Building Models

Exploration /
Visualization



- Exploratory analysis
- Plotting and graphing
- Dashboard creation

WHAT TOOLS DO DATA SCIENTISTS USE?

EXAMPLE PROJECT



<http://bit.ly/pythonsnow>

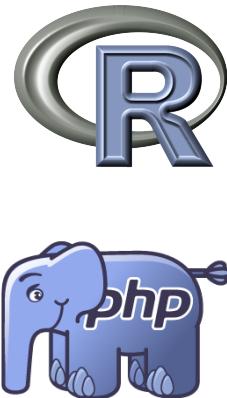
PYTHON 101

PROJECT DEMO

WHAT TOOLS DO DATA SCIENTISTS USE?

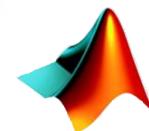
OTHER TOOLS

Querying / Collection



- Database Queries
- Web Scraping
- API Calls

Manipulation / Modeling



MATLAB



- Munging and wrangling
- Merging and enhancing
- Building Models

Exploration / Visualization



MicroStrategy®



Chart.js



- Exploratory analysis
- Plotting and graphing
- Dashboard creation

LEARNING OBJECTIVES

- Discover the history of Python and how it compares to other programming languages.
- Touch on fundamental Python programming techniques and tools.
- Discuss its applications in data science and the types of problems it can solve.
- See where Python programming fits into the data workflow.
- Apply your new skills to solve a real-world problem with Python.

PYTHON 101

CODING

PYTHON 101

Q & A