

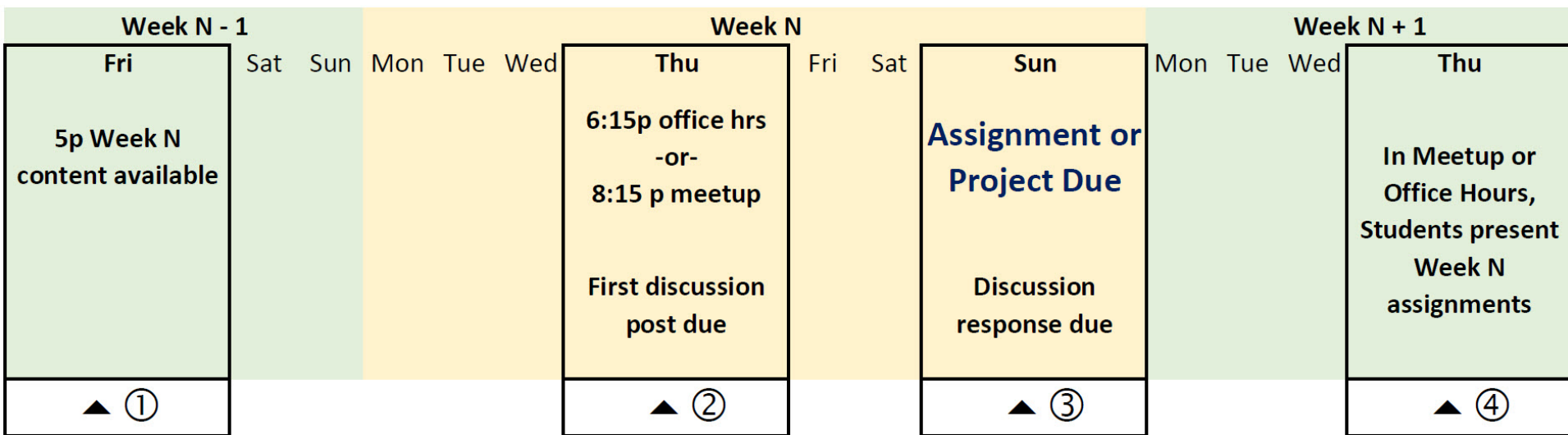


607 Meetup #1 Feb 4, 2016

Tonight:

- IS 607: Getting the work done
- Where does **loading data** and **preparing data** fit in data science?: *A look at types of analysis, roles, workflows, and skills.*
- Week 2 assignment: some sample code
- Drill down: R factors

Assignment Workflow:



Course Policies:

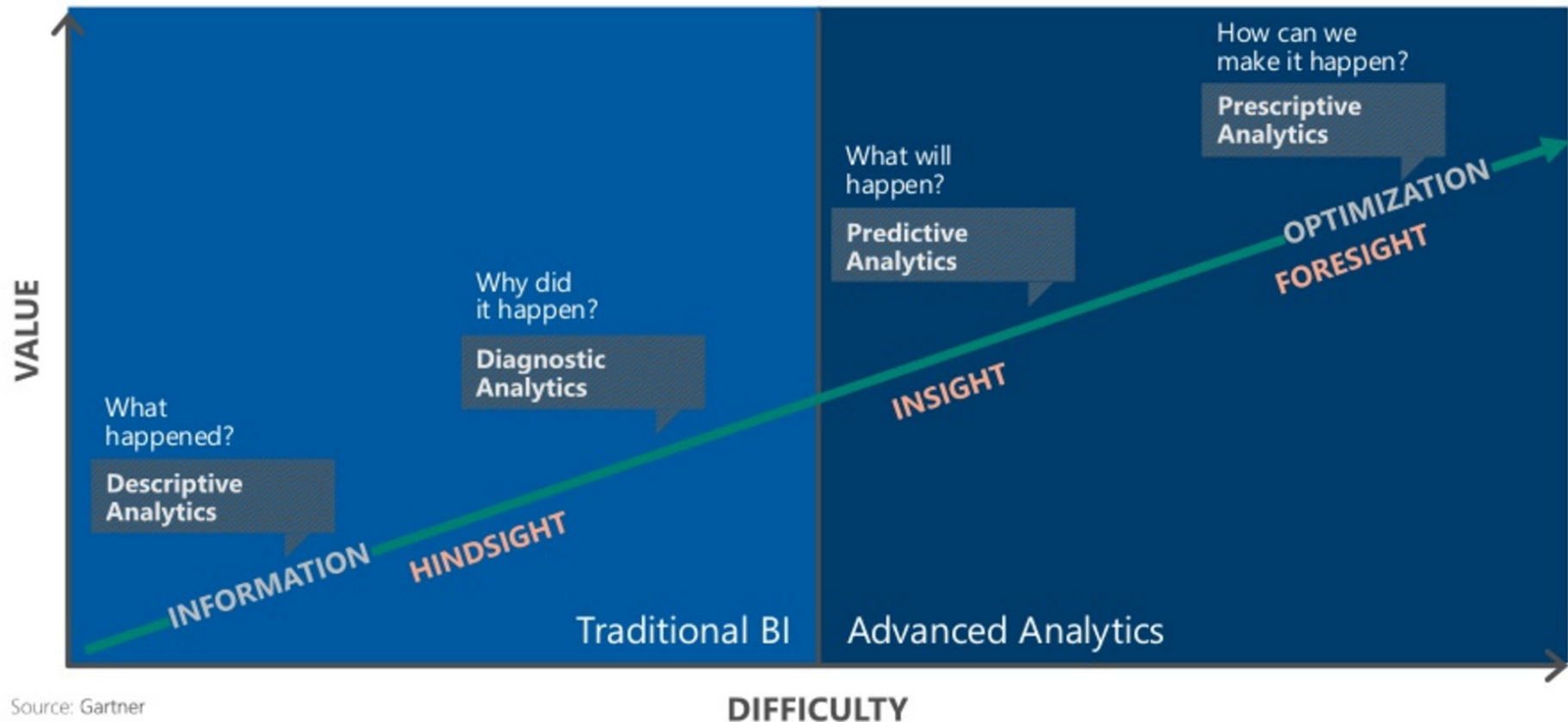
- Working Together
- Working with Found Code

What is the difference between data analysis and data science? What is the difference between structured data and unstructured data?

What is the difference between reporting and analysis?

Advanced Analytics

Beyond business intelligence



Next Bridge Workshop:

Summer 2016

SQL Programming

Data Science Math

R Programming

1

Programming & Data
Acquisition I
(R, unstructured data)

2

Probability and
Statistics

Programming & Data
Acquisition II (Python)

3

Computational
Mathematics

Visual Analytics

Mathematical
Modeling

4

Business Analytics
& Data Mining

5

Simulation and
Modeling

Predictive
Modeling

6

Linear Programming
and Optimization
[being developed]

Big Data and
Machine Learning

Capstone Project

Web Analytics

[network analysis and
text mining]

Financial
Engineering

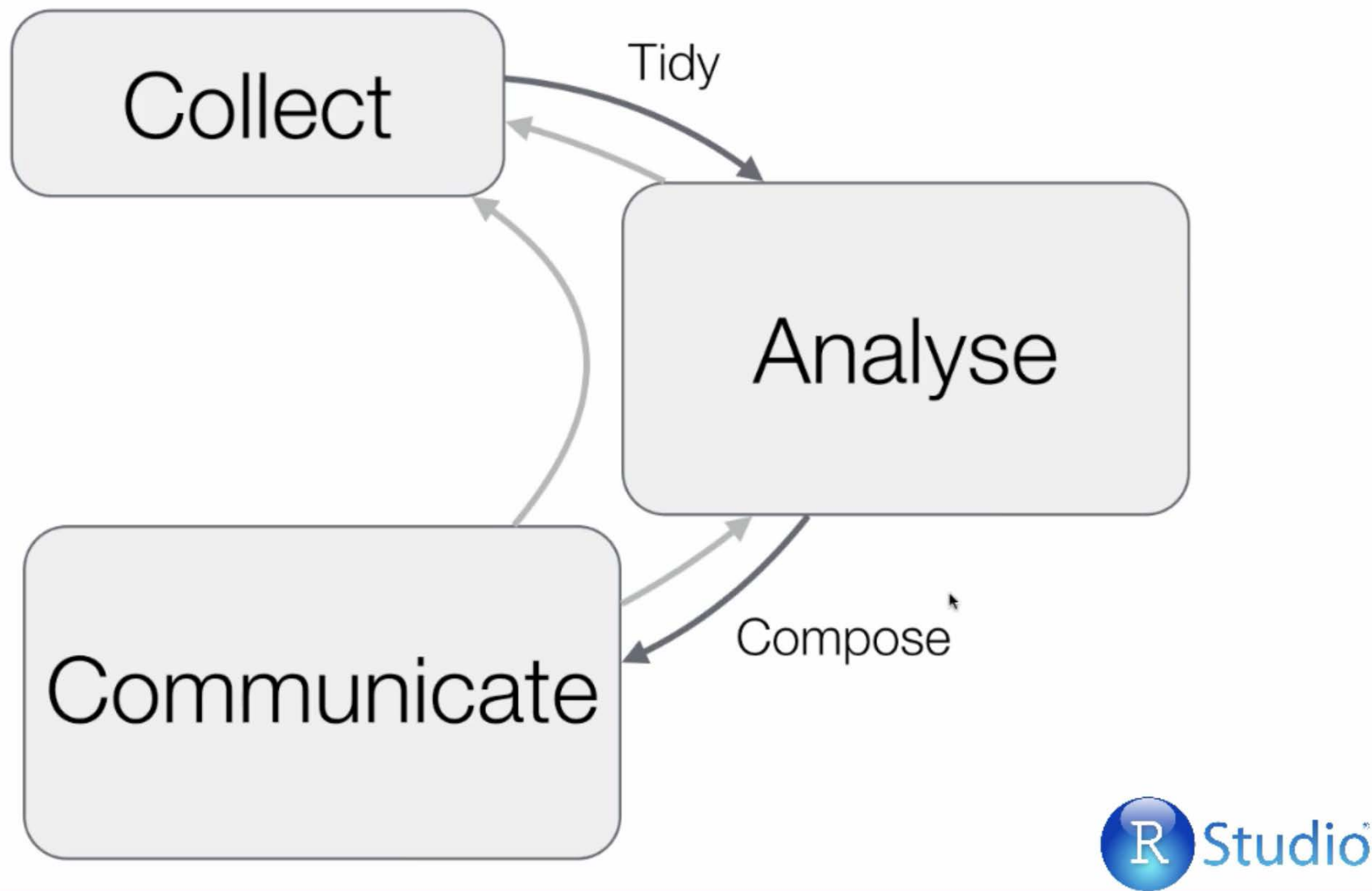
[special topics]

Recommender
Systems

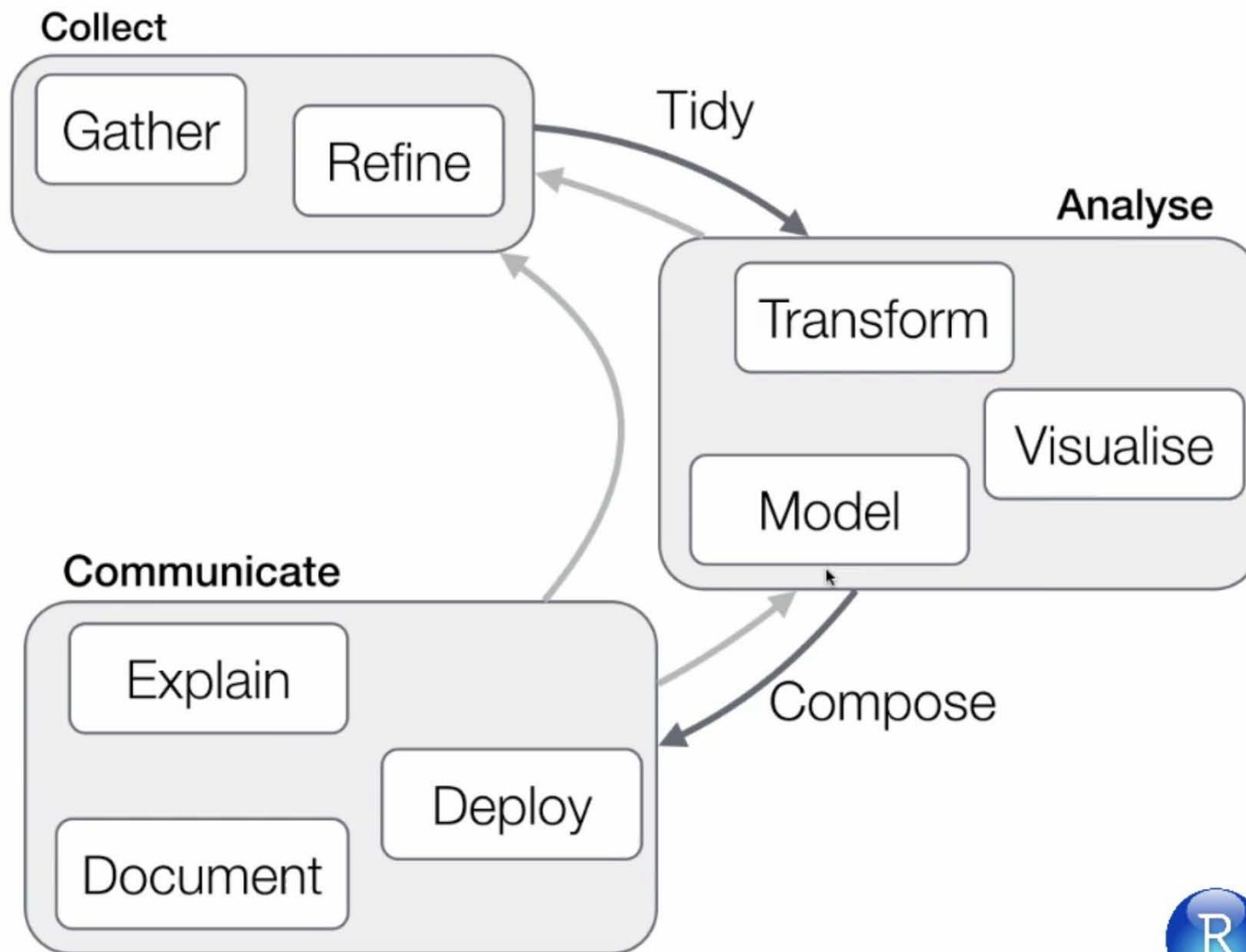
[special topics]

Independent
Study

What does a data analysis workflow look like?



Presentation by Hadley Wickham on “The Grammar and Graphics of Data Science”,
<https://www.rstudio.com/resources/webinars/archives/>



Presentation by Hadley Wickham on “The Grammar and Graphics of Data Science”,
<https://www.rstudio.com/resources/webinars/archives/>

OSEMN

Obtain

Scrub

Explore

Model

iNterpret

<http://www.dataists.com/tag/osemn/>

What are the roles in a data analysis project? What is the difference between coding and scripting?

What are the roles in a data analysis project?

Business Person (stakeholder/analyst/value architect/customer)

report gopher / data analyst / **Data Scientist**

Software engineer (implementer, operationalizer, automates business process)

...explores, builds models

Data
Scientist

Domain
Expert

*...frames, sets
context*

Software
Engineer

*...operationalizes
(implements models
in algorithms)*

R, **Python**, SQL, etc.
...explores, builds models

Java, C++,
Python,
SQL, etc.

Data Scientist

Domain
Expert

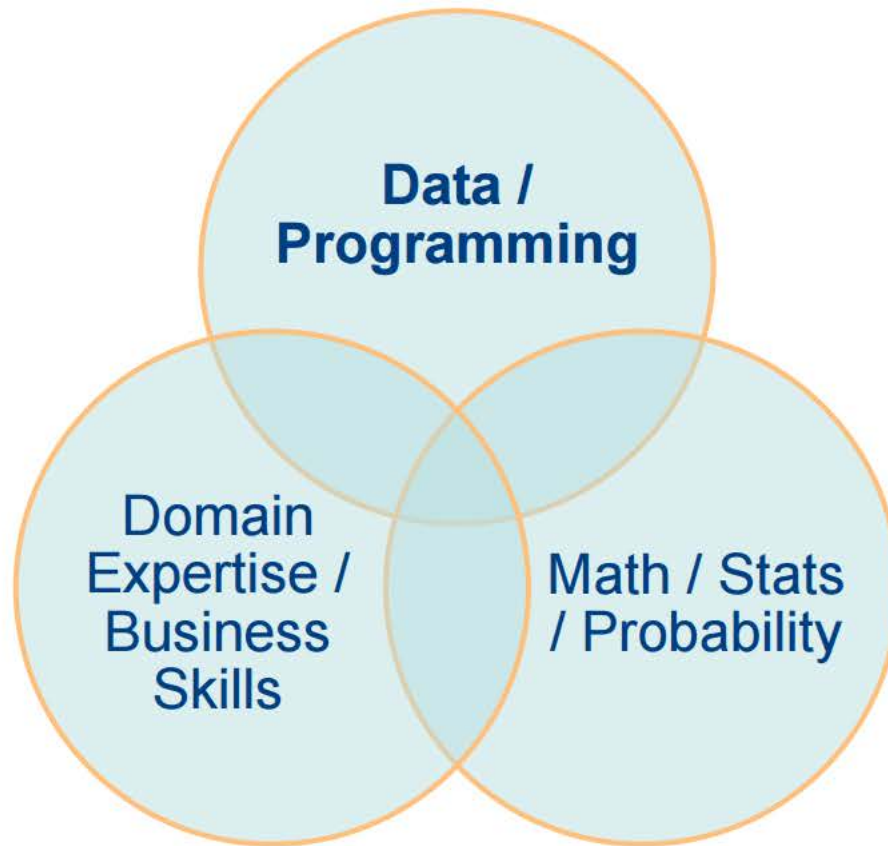
*...frames, sets
context*

Excel, Tableau,
SQL, etc.

Software
Engineer

*...operationalizes
(implements models
in algorithms)*

Key Skills?



Source:based on Drew Conway

Key Skills?

Source: Analyzing the Analyzers, O'Reilly 2013,
<http://www.oreilly.com/data/free/analyzing-the-analyzers.csp>

Business	ML / Big Data	Math / OR	Programming	Statistics
Product Development	Unstructured Data	Optimization	Systems Administration	Visualization
Business	Structured Data	Math	Back End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian / Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

What is an R Studio markdown file (rpubs.com)? What are its benefits?

Two good sample mushroom data projects:

Ken Marcus:

https://github.com/kennygfm/MSDA_Week1/blob/master/Week3AssignmentKenMarkus.Rmd

Vithalani Chirag:

<http://rpubs.com/chirag/126116>

R Factors

<http://rpubs.com/catlin/150599>