

Handling Big Data



Matthew Renze

DATA SCIENCE CONSULTANT

@matthewrenze www.matthewrenze.com



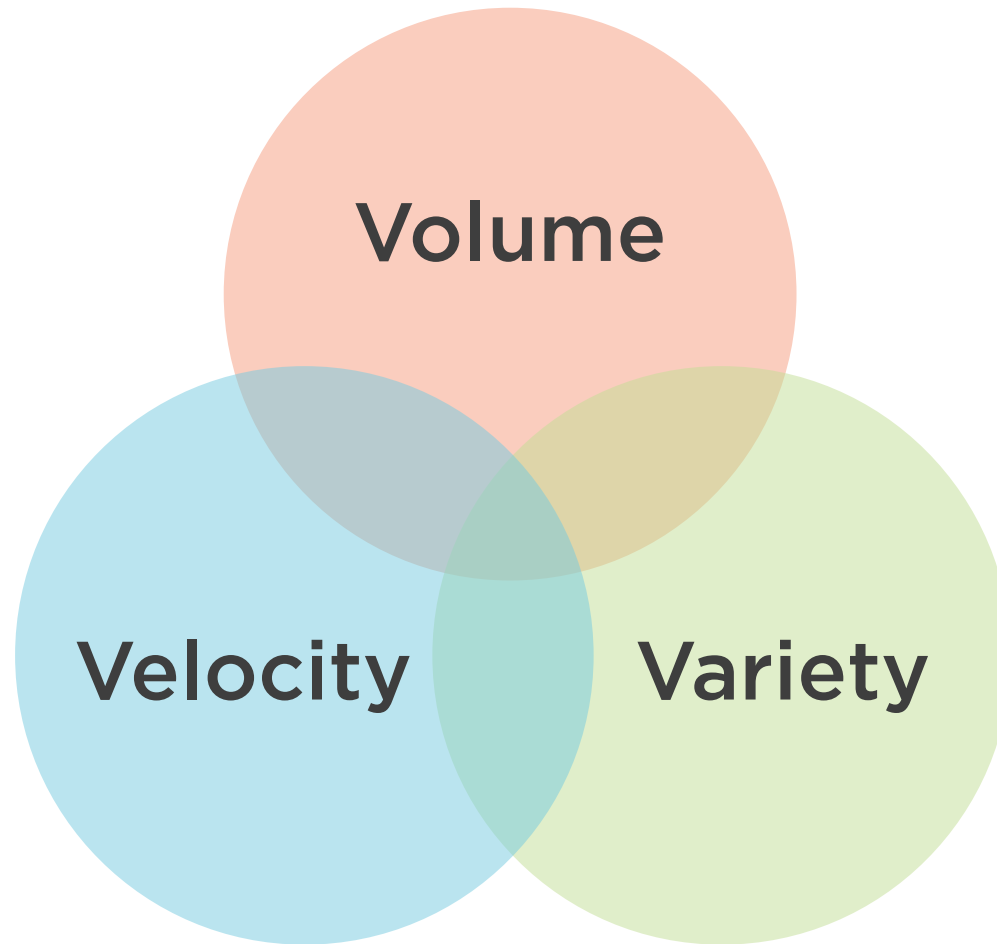
Overview



Handling Big Data
Demo



What Is Big Data?



Big Data Is a Moving Target

What is Big Data today...
Will not be Big Data tomorrow



Do You Have Big Data?

Can it fit in memory?

Can it fit on your hard drive?

Can you process it in a day?

Does it fit into tables?



Big Data Decision Table

Big Data Decision Table				
Class	Size	Rows	Storage	Management
Small	Gigabytes	Millions	Memory	R with desktop computer
Medium	Terabytes	Billions	Hard disk	R with medium-data extensions
Big	Petabytes	Trillions	Clusters	R with big-data extensions



Handling Big Data in R

More hardware

Subsetting

Sampling

Microsoft R Open



Handling Big Data in R

Medium data packages

Big data packages

Microsoft R Server







1. How are petal length and petal width related?
2. Can we predict new petal width values?

Load the Data



Inspect the FFDF Class



Create a Big Linear Model



Inspect the model



Create Scatterplot

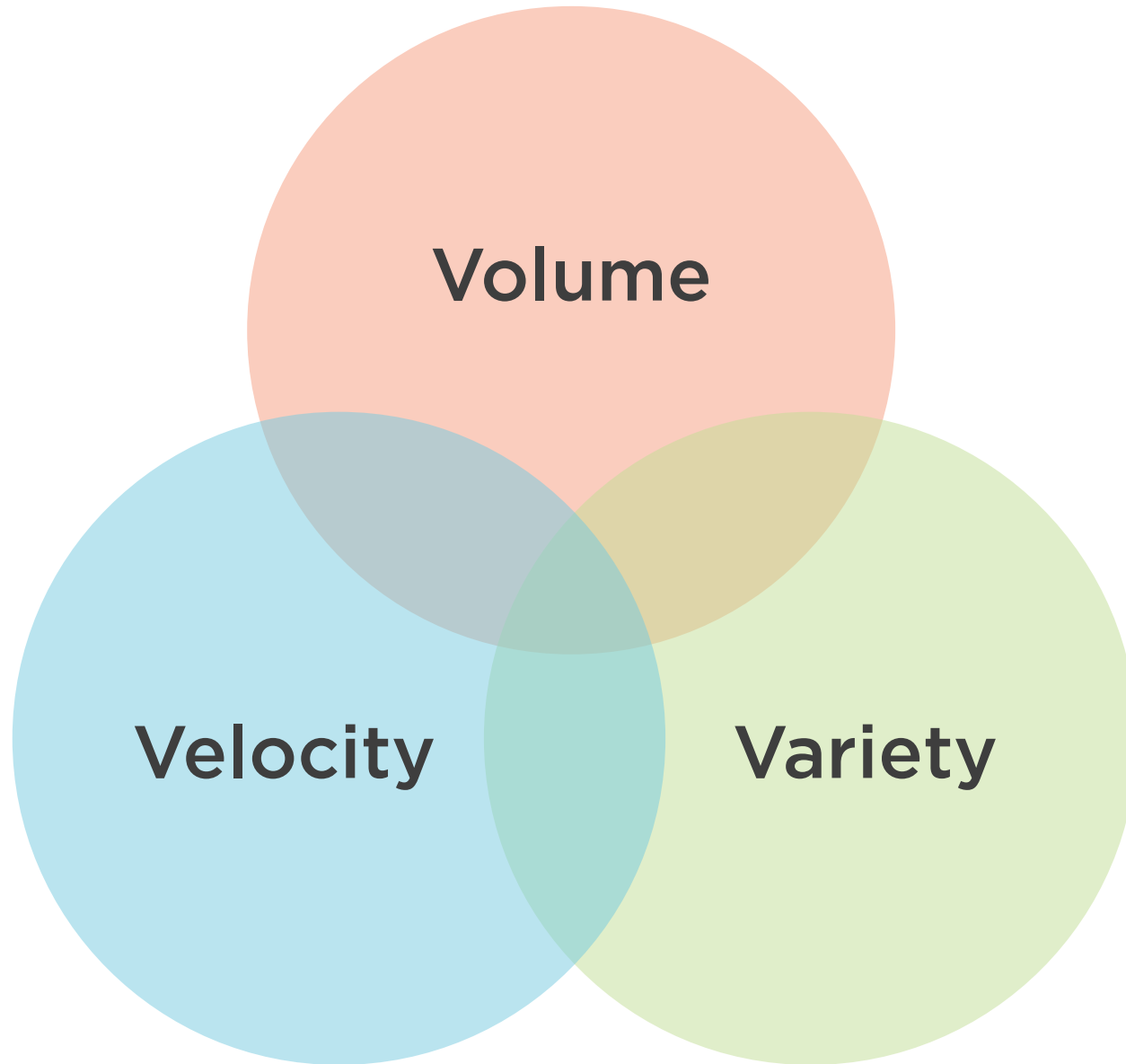


Draw the Regression Line



Predict New Values





Summary



Handling Big Data

Demo

