Organising, Visualising, and Describing Data

Data Types and Data Organisation

© 2021 by Keith Tan. All rights reserved.

No part of this video/slides may be reproduced or transmitted in any form or by any means, electronic or mechanical, without the written permission of the copyright holder.

PREP nuggets

Organising, Visualising, and Describing Data

Data Types and Data Organisation

1. Numerical vs Categorical
2. Cross-sectional vs Time Series
3. Structured vs Unstructured

PREP nuggets

Numerical vs Categorical Data

"Quantitative Data"

Numerical Data

values that can be counted or measured

Discrete

Countable

(***finite*** *number of values)*

Continuous

Any fractional value

(***infinite*** *number of possible values)*

*values do **not** have to stretch to infinity*

Numerical vs Categorical Data

"Quantitative Data"

Numerical Data

values that can be counted or measured

Discrete

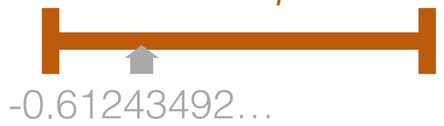
Countable

*(**finite** number of values)*

Continuous

Any fractional value

*(**infinite** number of possible values)*



Range-bound

(e.g. correlation [-1,1])

Numerical vs Categorical Data

Numerical Data

values that can be counted or measured

Discrete

Continuous

"Qualitative Data"

Categorical Data

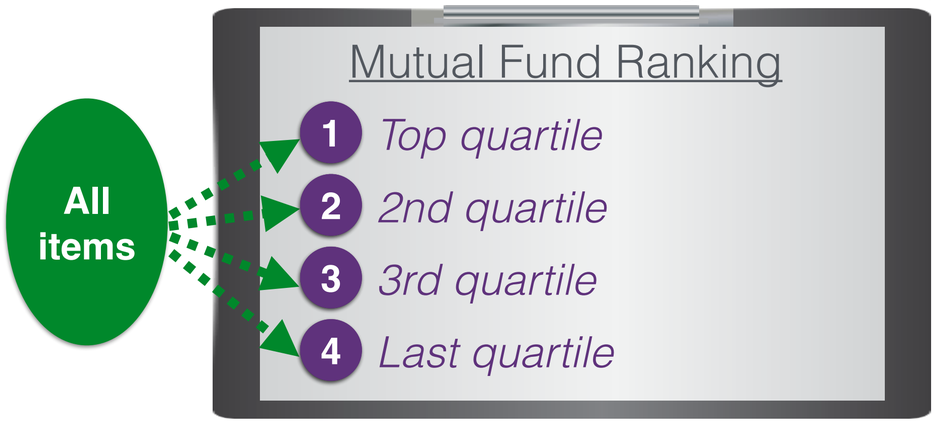
labels used to classify a set of data into groups

Nominal

*(labels have **no logical order**)*

Ordinal

*(can be **ranked in logical order**)*



PREP nuggets

Data Types and Data Organisation

1. Numerical vs Categorical 2. Cross-sectional vs Time Series 3. Structured vs Unstructured

Numerical vs Categorical Data

Numerical Data

values that can be counted or measured

Discrete

Return of Mutual Fund — **Continuous**

Difference is measurable

Fund B: 6% — **Fund A: 10%**

"Fund A outperforms Fund B by 4% pts"

Categorical Data

labels used to classify a set of data into groups

Nominal

*(labels have **no logical order**)*

Industry Classification

- *Manufacturing*
- *Retail*
- *Food & Beverage*

Ordinal

*(can be **ranked in logical order**)*

Mutual Fund Ranking

1. *Top quartile* — **Fund A:Q1**
2. *2nd quartile* — **Fund B:Q2**
3. *3rd quartile*
4. *Last quartile*

Difference is NOT measurable

"Fund A performed better than Fund B"

PREP nuggets

Data Types and Data Organisation

1. Numerical vs Categorical
2. Cross-sectional vs Time Series
3. Structured vs Unstructured

Cross-sectional vs Time Series Data

Cross-Sectional Data

*comparable observations all taken at **specific time***

e.g. Sales data of 4 companies in 2011

Company	2011
A	$24,000
B	$122,500
C	$34,000
D	$72,800

Cross-sectional vs Time Series Data

Cross-Sectional Data

*comparable observations all taken at **specific time***

Time Series Data

*set of observations taken periodically over a **period of time***

Company	2011	2012	2013	2014	2015
A	$24,000	$27,000	$31,000	$38,500	$45,900
B	$122,500				
C	$34,000				
D	$72,800				

1D Array

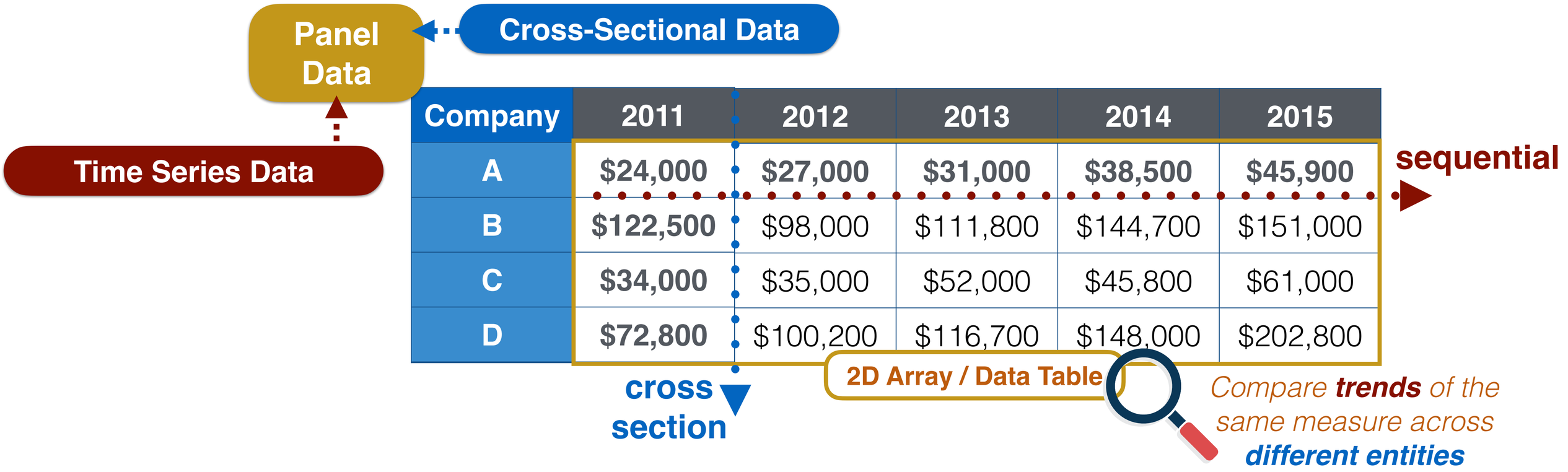
Represent single variable

e.g. annual sales for Company A

Forecasts

Identify trends, cycles, patterns

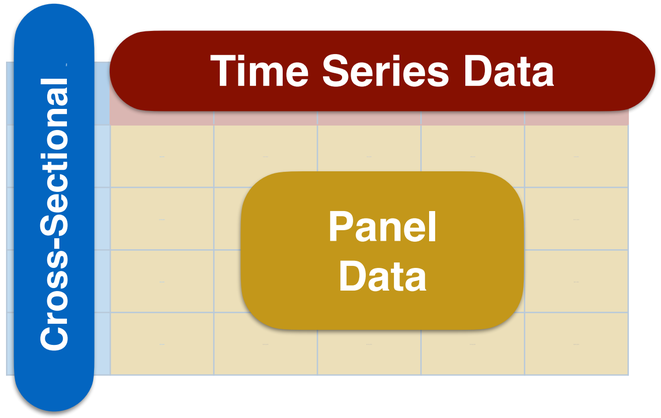
Cross-sectional vs Time Series Data



Structured vs Unstructured Data

Structured Data

*organised in a **defined** way*



Market data
e.g. security prices

Analytical data
e.g. earnings forecasts

Fundamental data
e.g. accounting values

Unstructured Data

*information in a form with **no defined structure***

Generated by:

Investor sentiment

Business sentiment

Business activity

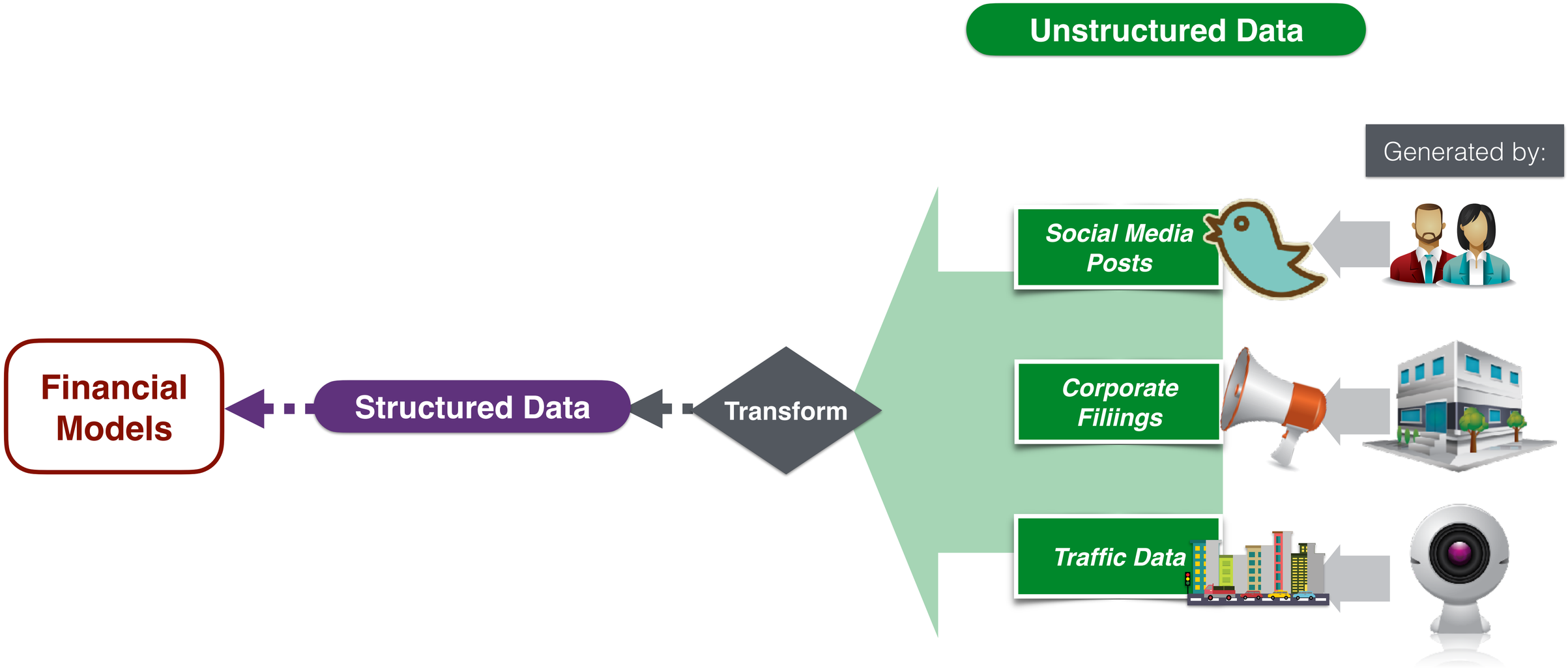
Challenge to process and analyse large amount of data

PREP nuggets

Data Types and Data Organisation

1. Numerical vs Categorical
2. Cross-sectional vs Time Series
3. Structured vs Unstructured

Structured vs Unstructured Data



prepnuggets.com