

# DataDwarf

Sean Spearman | Cody Brown  
Ray Smets | Aimee Galang | Tim Shen

## Problem

- ◆ Discerning what kind of relationships exist among raw data is difficult.
- ◆ The process of sifting through large amounts of data and calculating relationships is tedious.
- ◆ Understanding the results of correlation calculations of data can be unintuitive.

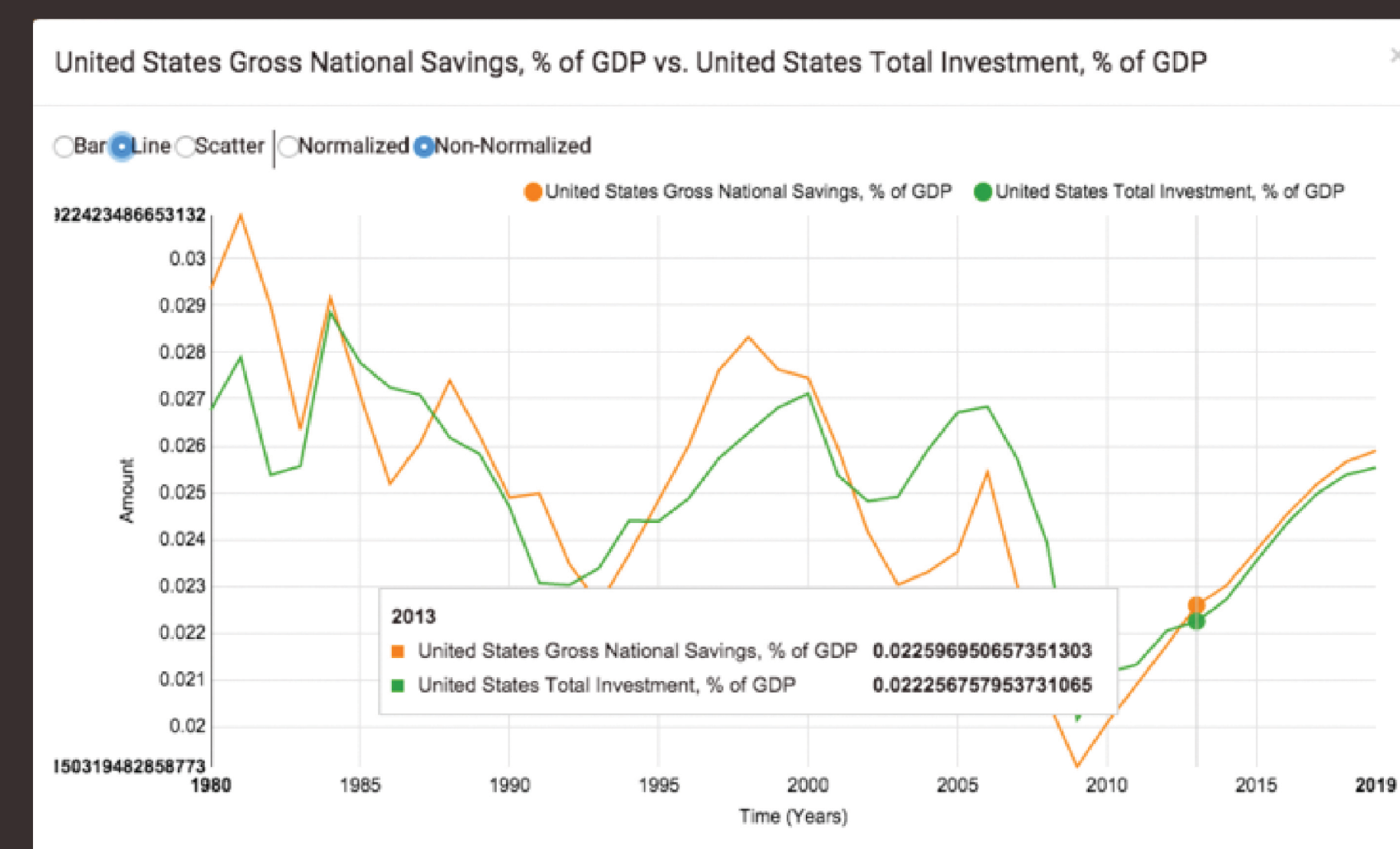
## Solution

- ◆ DataDwarf offers a platform for making the relationships between datasets obvious.
- ◆ Users can easily analyze data from the DataDwarf database with other data in the database or with their own data.
- ◆ Relationships are made obvious using the DataDwarf heatmap and can be further evaluated through more traditional graphs.

## Database Structure

- ◆ DataDwarf's data utilizes a tagging system that involves a many-to-many relationship between data sets and tags related to them.
- ◆ Tagged data allows for search and correlation by tag making relevant relations easy to find.

## Graph



- ◆ Once a relationship of interest has been found, the details of it can be evaluated using more traditional graphs, including bar charts, line graphs, and scatter in both normalized and non-normalized forms.

## Our Data + Your Data

- ◆ Upload your data to DataDwarf to leverage the power of visualization.
- ◆ Find relationships between your data sets and ours.
- ◆ Find relationships hidden within your data.

## Acknowledgements

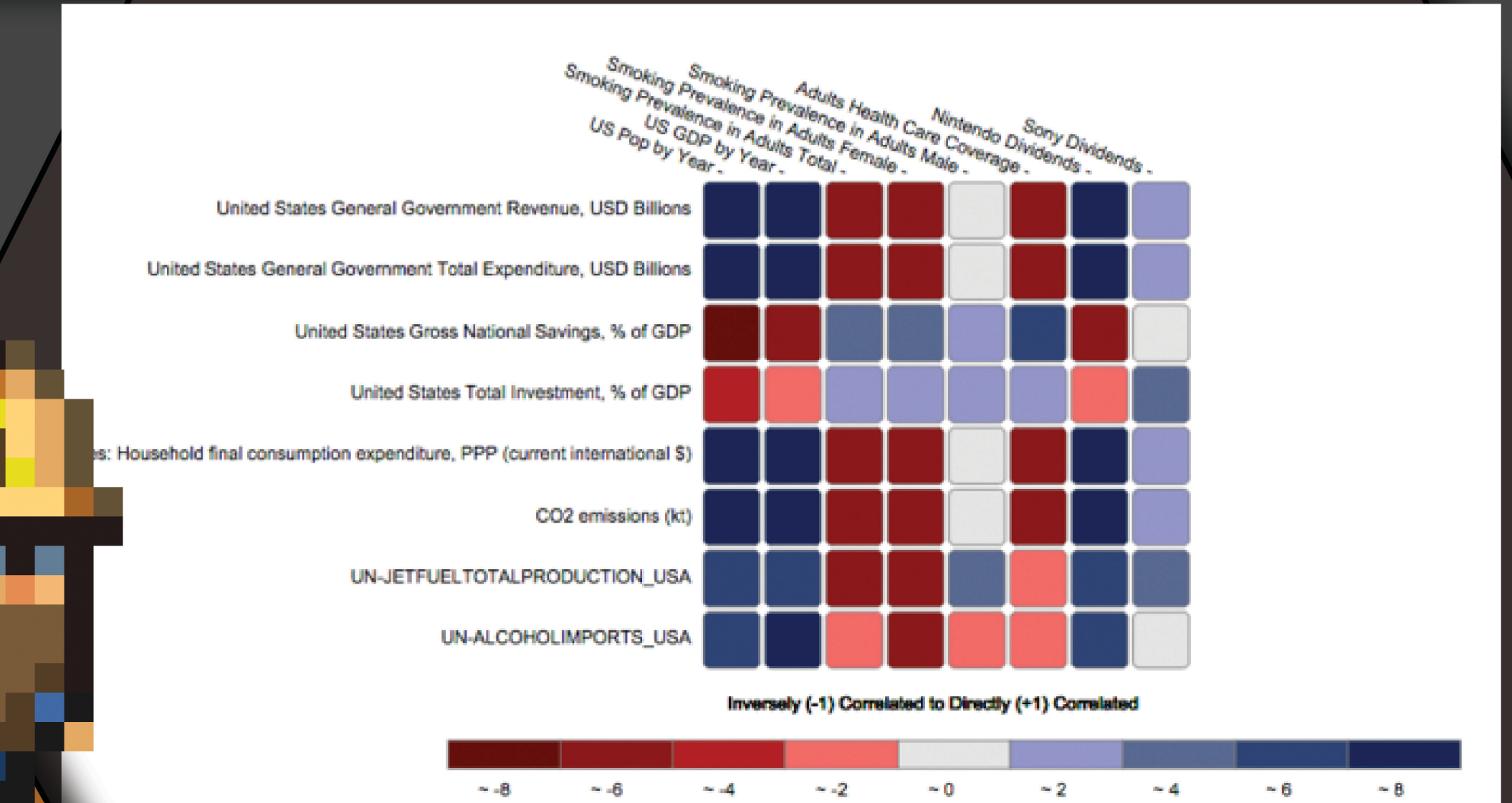


Victor Moreira, Sam Chapin, and the HG Data team



Chandra Krintz, Tim Sherwod, and Kyle Jorgensen

## Explore Heatmap



- ◆ The Explore heatmap visualizes the degree of correlation between many data sets using color.
- ◆ For example, the dataset in the first column is positively correlated with the data set in the first row. Conversely, the data set in the third column is negatively correlated with the data set in the first row.

## Explore

- ◆ If a user is unsure which data set is of interest to them, they can browse the DataDwarf database with the Explore feature.
- ◆ The Explore feature compares many data sets with many others by means of the Explore heatmap.

## Analyze

- ◆ If a user has a particular data set in mind, they can find data sets related to it with the Analyze feature.
- ◆ The Analyze feature compares one data set with many others by means of the Analyze heatmap

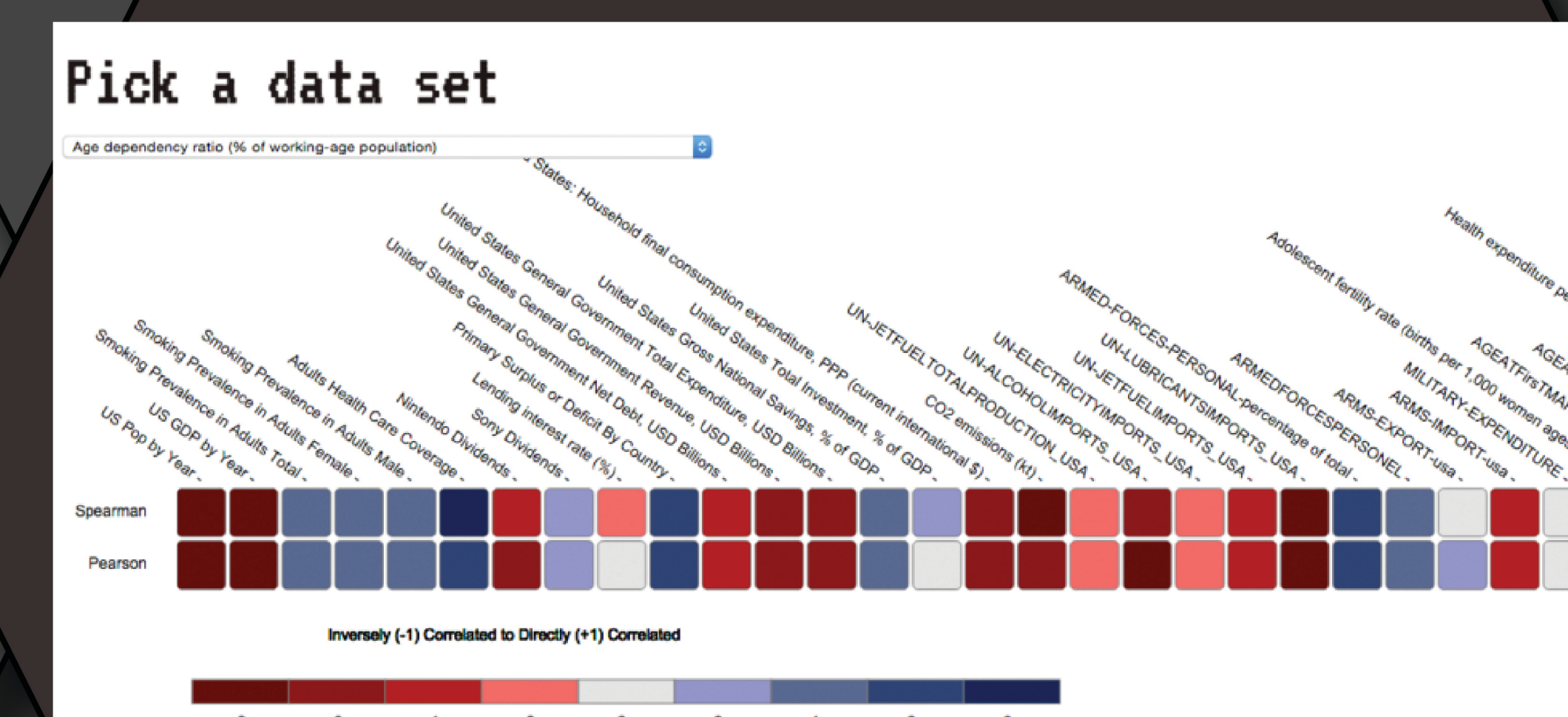
## Correlate

- ◆ A correlation is a mutual relationship between two sets. Positive correlations indicate that the growth of one set is tied to the growth of the other while negative correlations indicate that the growth of one set is tied to the decline of another.
- ◆ DataDwarf uses a multitude of correlation measures to track how closely two data sets are related and make relationships obvious.

## Technologies Employed



## Analyze Heatmap



- ◆ The Analyze heatmap visualizes the degree of correlation between one data set and many others using multiple correlation coefficients.
- ◆ Each row shows correlations using a different correlation algorithm while each column corresponds to the different data sets.