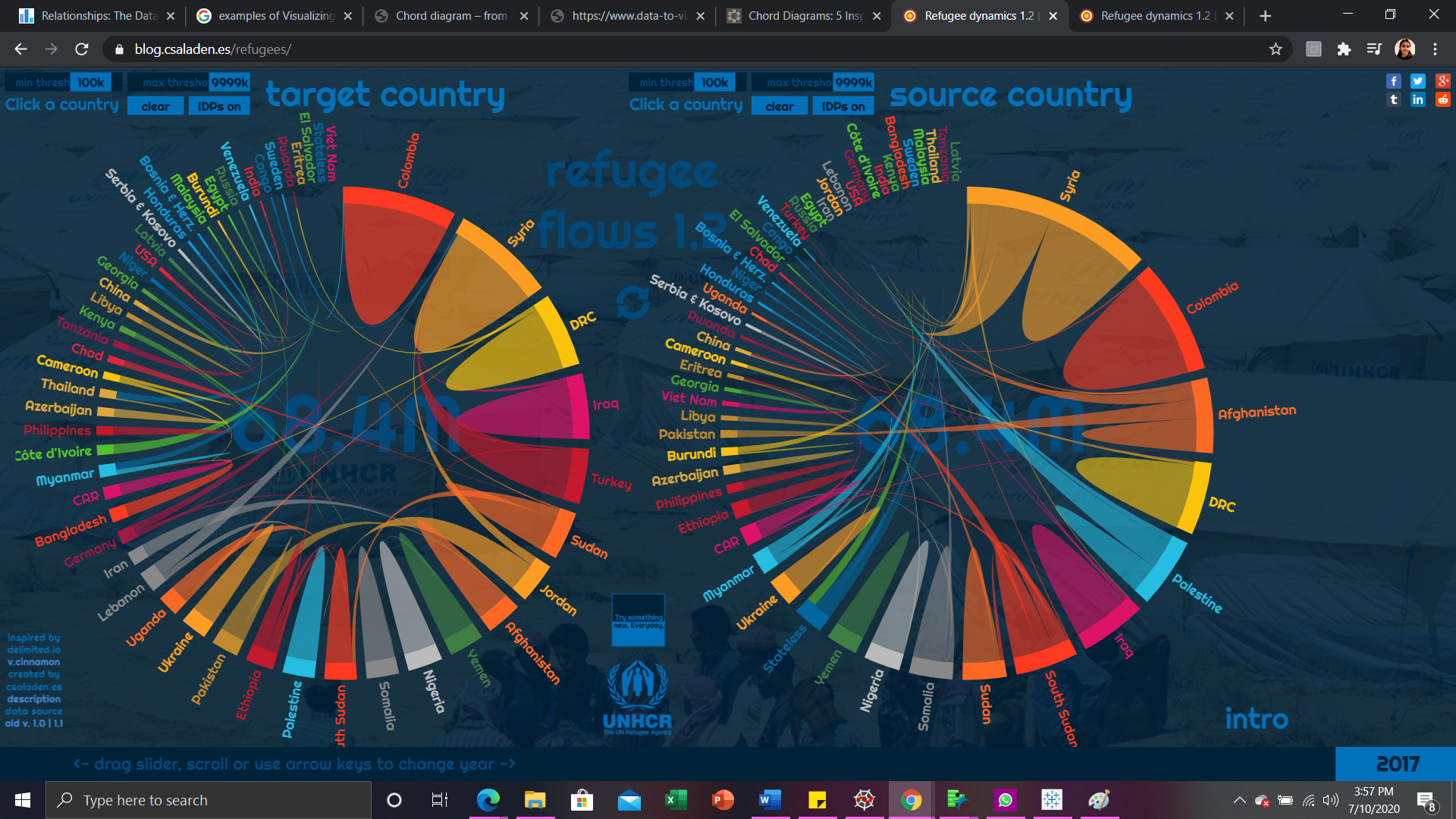
### **1. PART I: Data Visualization with Graphs**

A chord visualization will represent the connection or flows between numerous entities which are called nodes. Each of these entities is represented by a fragment on the outer part of the circular layout. Arcs are drawn between each of the entities and their size is proportional to the importance of the flow. The connection between entities makes chord visualizations ideal for comparing all the similarities between different groups of data or within a dataset.

We can hover on a specific group to highlight all of its connections. And for many other reasons interactivity is a real plus to make the chord diagram understandable. They are eye catching and quite popular in data visualization. We can even visualize weighted relationships between several entities.

The chord visualization below shows the domestic and global flows of refugees and the internally displaced persons from 1950 to 2017. It is actually vast but this chord visualization simplifies the interconnected relationships.



It maps the refugee flows between source and target countries during the period 1951-2017. It has two chord diagrams side-by-side, one of them is target-based, while the other one is a source based accounting. They display the receiving or the sending countries. We can filter out some countries from the visualization by clicking on their label. We can even filter out all countries with flows smaller than a threshold (set to 100000 people by default). All the data is dynamic and to change the year we can drag the time slicer or scroll with the mouse or use the arrow keys.

It’s quite a powerful tool for analyzing the refugee flows of the past half century. The total number of refugees is displayed at the center of the chord diagrams. We can set a floor and ceiling for the displayed flows, with which we can visualize like only refugee flows that fall between 1500 and 2500 people.

Briefly, insights till the year 2017 are,

* The world’s refugees total 68.4 million when including IDPs, 21.7 without.
* In 2017, Colombia is the largest source of refugees, by far 7.6 million people in total.
* Syria, DRC and Iraq have a huge number of IDPs (6.8,4.6, and 4.1 million respectively)
* Without IDPs, Turkey is the largest host country, hosting 3.7 million refugees, Iraq - 152 thousand, Afghanistan - 163 thousand, and Syria - 2.1 million.
* Taking into account IDPs further dwarfes the flows targeted towards developed countries. Decreasing the filters by 100 to 1000 times and displaying only the flows between 2000 and 10000 refugees, the European and American flows are highlighted.
* We can see that the developed world welcomes refugees almost equally from a number of source countries, which is a different pattern from those of the developing countries – many next to the conflict-affected areas – who mainly host refugees from one or two countries – but the scale is about 10-100 times smaller.

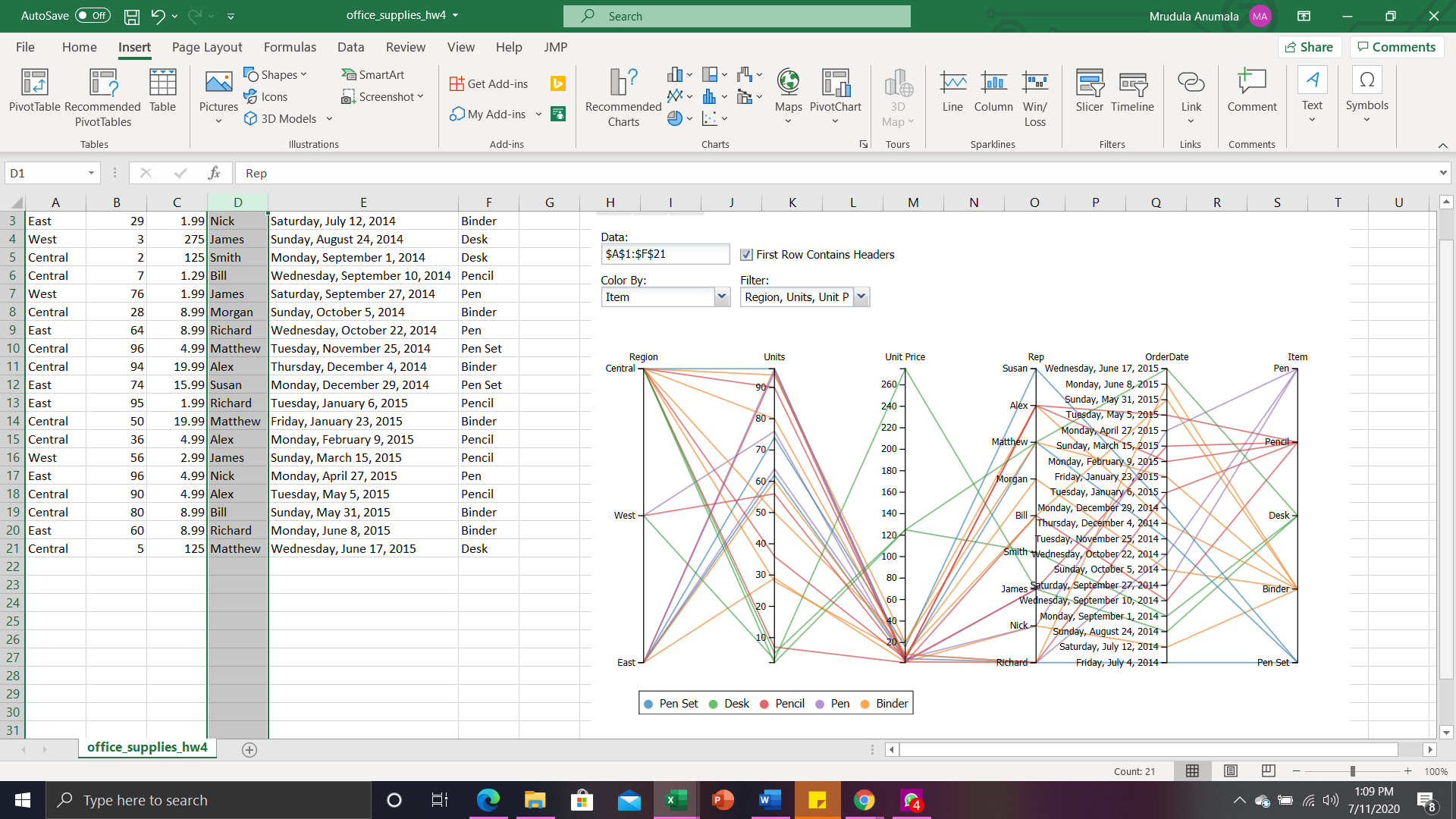
### 

### **2. PART II: Multivariate Data Visualization with Geometric Representation**

Below visualization is a Parallel coordinate plot. Each of the attributes is represented with a vertical axis in this plot. In a 2D space, *n* attributes are arranged as *n* uniformly spaced vertical lines. A data record with *n* attributes is manifested as a connected set of *n* points, one on each of the axes. Attributes can be normalized if required, so that their minimum and maximum values correspond to the bottom and top points.

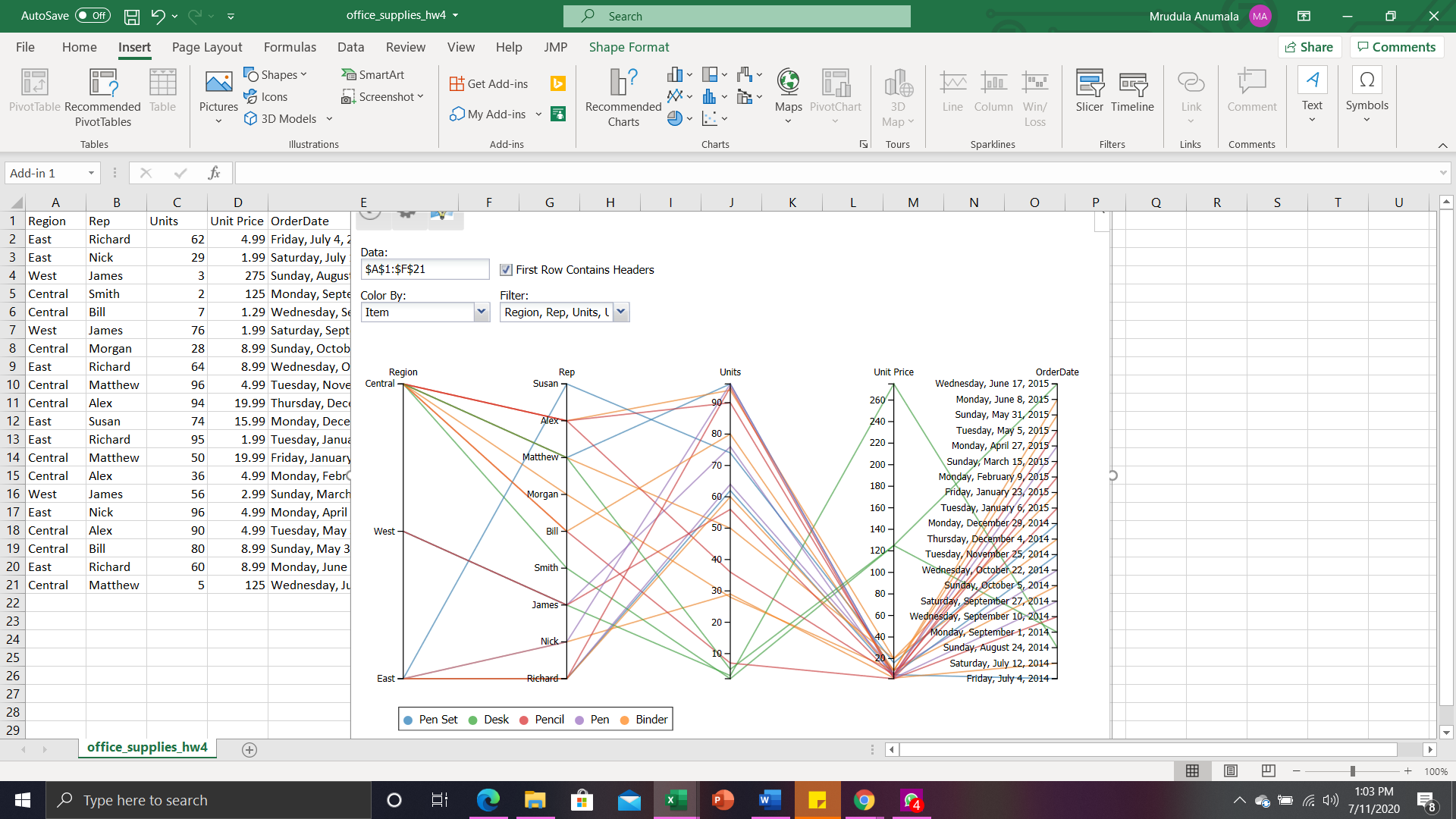
This visualization is effective in revealing the relationships between adjacent axis attributes. It is effective in showing the distributions of attributes as well. And based on the order of the axes, the effectiveness of a visualization is greatly impacted. When the number of data records is high relative to the display or if there are many data records having the similar or same values, overlapping of line segments will occur. To address any of the problems, we can apply interaction techniques like selecting only a subset of the data for visualization, changing the order of the axes.

When coupled with appropriate interaction techniques, these plots can handle large datasets, but still we may have record overlapping or visual cluttering occurred for very large datasets. It can handle medium and high dimensional data as well. All of the attributes are treated the same, but their order can affect what can be perceived. These visualizations are very effective for detecting outliers and correlation among different variables.



Below insights can be drawn from the parallel coordinate plot shown above,

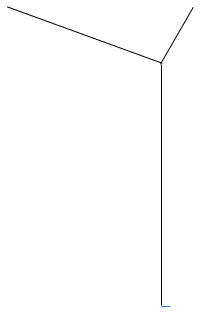
* Most of the order supplies are from the Central region, then from the East region, and atlast from the West region.
* West region has only three of the five items supplied, and they include - Pen, Binder, and Desk.
* Looking at the Units axis, it is clear that the item Desk has the lowest number of units which is less than 10, supplied for all the regions.
* And there is a negative relationship between the Units and Unit price attributes. All the items which have a higher number of units supplied, have a low unit price ranging between 1 and 20. Whereas, on the other hand, for the item Desk in particular, it has a lower number of units supplied, and has a comparatively high Unit price of 275 and 125.
* Number of office supplies for each representative ranges from 1 to 4. Richard had 4; James, Matthew, and Alex had 3; Nick and Bill had 2; Smith, Morgan, and Susan had 1.



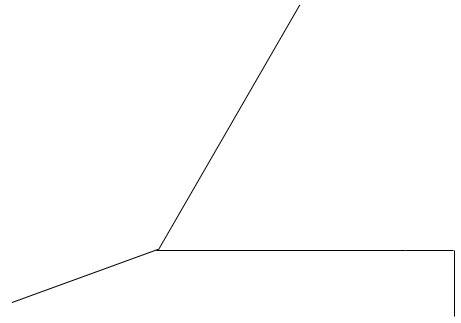
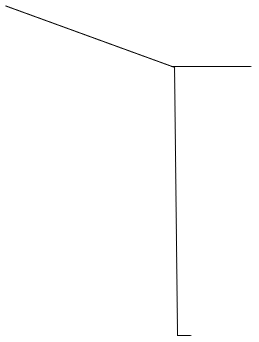
On rearranging the order of attributes, we can see the regions of each of the representatives. James is the only one from the West region. Susan, Nick and Richard are from the East region. And the rest of the representatives Alex, Matthew, Morgan, Bill, and Smith come under the Central region.

### **3. PART III: Multivariate Data Visualization with Icon Representation**

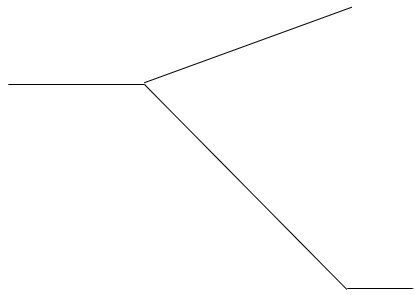
Stick figure is one of the classical icon-based techniques where we will map two attributes to the display axes and the rest of them to the length, rotation angle, color or thickness of the limbs as represented for the iris data set below. With respect to the display dimensions, when we have relatively dense data items, packed icons will show some texture patterns which will vary according to the features of data and these can be detected by preattentive perception. But the selection of appropriate graphical attributes is something that the visual discernment of an important pattern is mostly dependent on. And therefore, the selection process is deterministic and can be bottleneck.

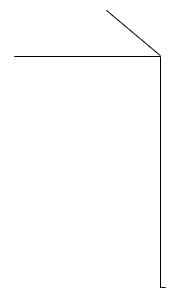
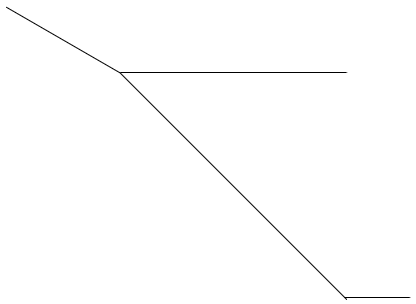
ID = 1 ID = 2

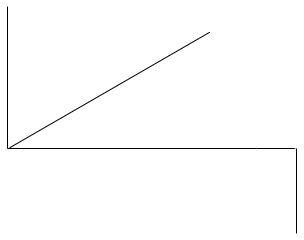
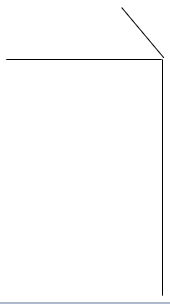
ID = 3 ID = 4

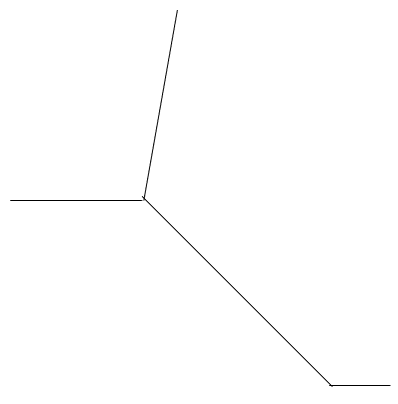
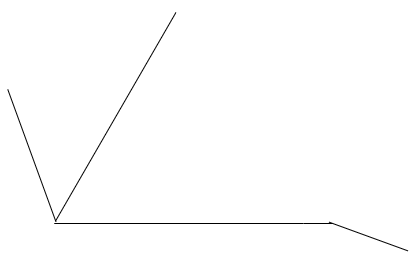
ID = 5 ID = 6

ID = 7 ID = 8

ID = 9 ID = 10

ID = 11 ID = 12

In the given IRIS data set, apart from the ID attribute which shows us the indices of the records, we have 5 other attributes SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, and Species which are actually represented in the stick figures. We have 12 records in our data set, and each record is represented with one stick figure.

We have 3 different species : Iris-setosa, Iris-versicolor, Iris-virginica. This species attribute is represented in the stick figure by using the angle parameter. One among the three species is identified by looking at the inclination of the central line. Records with Iris-setosa as species are having their central line inclined at -90°, records with Iris-versicolor as species are having their central line inclined at -45°, and finally the records with Iris-virginica as species are having their central line inclined at 0°.

Along with this one nominal variable species, we have represented the remaining four quantitative variables as the limbs of the stick figure.

### **4. PART IV: Multivariate Data Visualization with Pixel-Based Representation**

The idea of Pixel-Oriented representations is to encode each data unit in pixels based on some color scale. The color mapping of the pixels, arrangement of pixels in the subwindows and shape of the subwindows depend on the data characteristics and visualization tasks.

These techniques are further divided into two sub-groups,

* Query-independent: Entire dataset is visualized in these techniques (Space-filling curves, Recursive pattern technique).
* Query-dependent: Subset of data that is relevant to the context of a specific user is visualized in these techniques (Spiral technique, Circle segment).

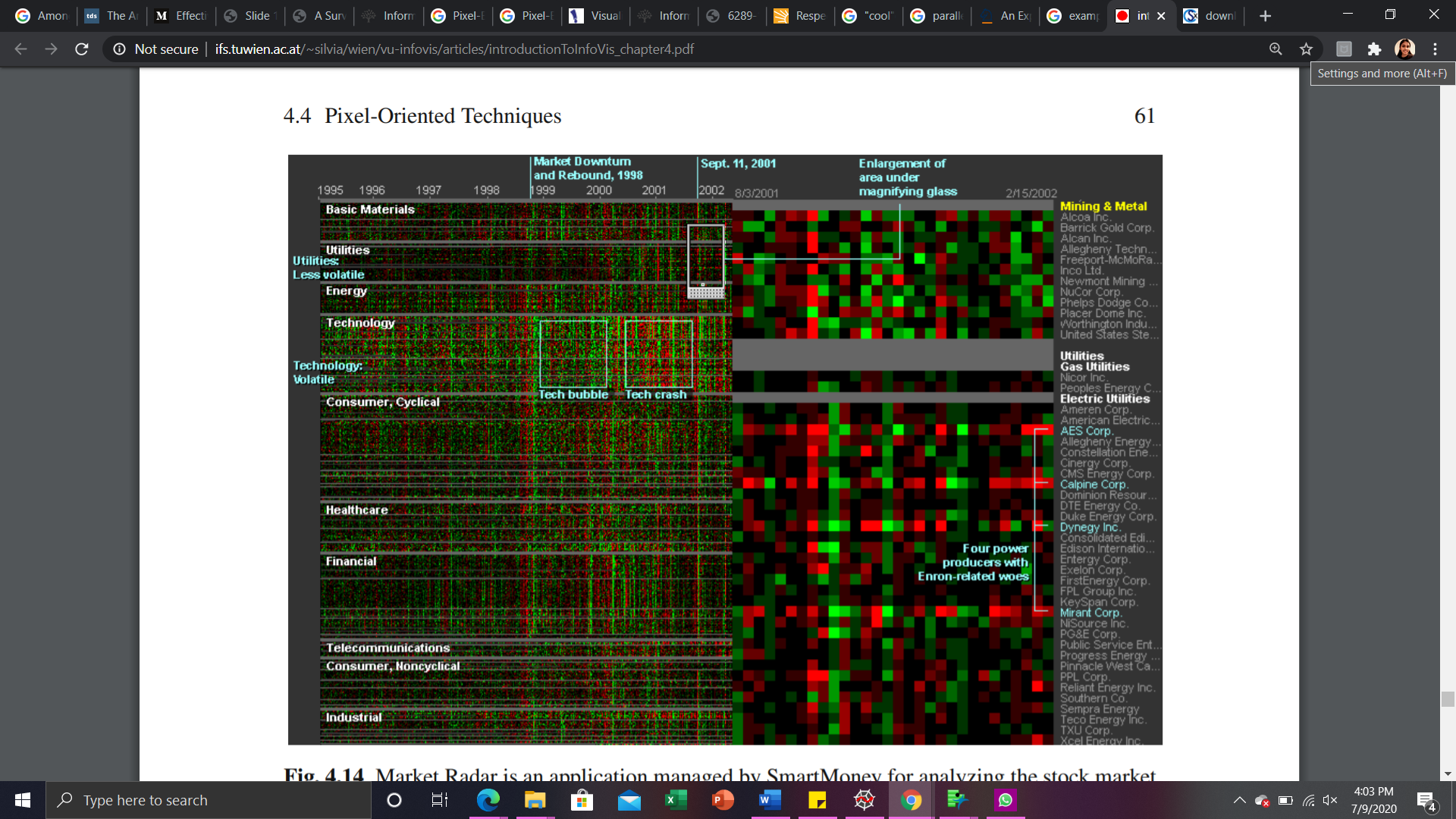
We can achieve effective visualizations for multivariate data by taking a series of considerations before using this technique, like,

* Shape of the window: What will be the shape of the window in which the pixels are displayed? Rectangular shape is usually used to represent, other shapes are proposed, but this remains to be the most suitable one.
* Visual mapping: What does each pixel represent in the visual mapping?
* Pixels arrangement: How will be the arrangement of pixels within each window?
* Color mapping: What color mapping will be adapted to the pixels?
* Windows ordering: What order will different windows on the physical space of the screen be arranged in?

The corresponding pixels appear at the same position in each respective window. With suitable rearrangement, users may observe the inter-relationships between attributes, trends and patterns in the underlying data. They are very effective in visualizing hierarchical data (which is also their limitation).

On summary, Pixel-Based techniques,

* Can fairly manage high- and medium- dimensional datasets
* Can manage very large datasets on the high-resolution displays
* Visual cluttering and data record overlapping will not occur as each data record is uniquely mapped to a pixel
* As color is not an effective parameter in visualizing quantitative variables, the level of revealing the quantitative relationships between variables is limited.



This interesting example is Market Radar, history of the shares quoted on the U.S. stock exchange can be analyzed. Over a substantially long period of time, this will allow us to examine the entire market dynamic at a glance because, over the last 8 years, the price variation of 500 shares is visualized. The weekly price variations are taken into account in the Market Radar. Each week is illustrated in the left part of the chart by a column, where a different row of the matrix corresponds to each stock price. In order to examine the progress of different types of industries or sectors, stocks are grouped in sectors. Each weekly price shift is illustrated by a dot in the matrix on the left. If the stock went down in that week, it is represented by red, and by green if it went up. While the concentration of the color shows the deviation with respect to the past week. We can clearly discover that the shares from the technology sector have the highest instability. The example displayed in the chart presents an intensely negative week for most of the stocks, corresponding to the attack on the Twin Towers in New York on September 11, 2001.

**References:**

[1] <http://people.stat.sc.edu/hansont/stat730/MultivariateDataVisualization.pdf>

[2] <https://www.saedsayad.com/docs/multivariate_visualization.pdf>

[3]<https://www.interaction-design.org/literature/article/information-visualization-an-introduction-to-multivariate-analysis>

[4] <https://www.data-to-viz.com/graph/chord.html>

[5] <http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/introductionToInfoVis_chapter4.pdf>