SMM635

Group 7

CASS Business School

Data Visualization Companion Document

Table of Contents

[1.Map of All Notebooks 2](#_Toc59060094)

[Data Cleaning Process: *Mapping* 2](#_Toc59060095)

[Visualizations: *Mapping* 4](#_Toc59060096)

[2.The Visualizations 5](#_Toc59060097)

[Geographical Visualizations 5](#_Toc59060098)

[Timeseries Visualizations 6](#_Toc59060099)

[Impactful Event Visualizations 7](#_Toc59060100)

[3.Impactful Event Overview 8](#_Toc59060101)

[Industries with the Most Impact: 8](#_Toc59060102)

[Events: 8](#_Toc59060103)

## 1.Map of All Notebooks

### Data Cleaning Process: *Mapping*

***cleaned\_data*** :

* Contains: NA
* Produces: clean\_data.csv
* Steps:
  + Remove space in columns
  + Select only london Based companies
  + Save to CSV

***get\_coordinate*** :

* Contains: cleaned\_data.csv
* Produces: tech\_roundabout.csv
* Steps:
  + Import cleaned\_data.csv to Data Frame (df)
  + Remove all na from data frame (df)
  + Select only East London companies in data frame (df)
  + Get longitude and latitude by requesting coordinates
  + Create a data frame called coordinates (coordinates)
  + Merge new Coordinate Data Frame (coordinates) with Original Data Frame (df)
  + Save as tech\_roundabout.csv

***Select Features:***

* Contains: cleaned\_data.csv
* Produces: feature\_data.csv
* Steps:
  + Import cleaned\_data.csv to data frame (df)
  + Select only columns that are necessary and add it to variable (df1)
  + Extract year and month from df1[‘IncorporationDate’]
  + Sort values in df1
  + Find all values greater than 1990 in df1
    - This was done because technology companies did not become prominent until the 90’s meaning we would not be able to visualize companies from every single year. This keeps the data clean and relevant
  + Save as feature\_data.csv
* ***Geospatial\_exploration***
* Contains:
* Produces:
* Steps:
  + Load geospatial data
  + Set coordinate reference system
  + Apply coordinate reference system to UK coordinates
  + Plot UK coordinates

***final\_data\_cleaning*:**

* Contains: tech\_roundabout.csv, features\_data.csv
* Produces: final\_cleaned.csv
* Steps:
  + Import tech\_roundabout.csv as tech\_roundabout\_full
  + Import features\_data.csv as features
  + Merge data on CompanyName as data frame called all\_data
  + Sort values by IncorporationDate in all\_data
  + Split sector code from SICCode.SICText into two separate columns in all\_data
    - This allows us to use the SIC Code to group companies and identify which sectors they belong to
  + Removed all unnecessary columns from all\_data
  + Removed all companies that are not currently active
    - Originally, the aim was to show companies dissolving or liquidating in our visualizations. However, we did not have any ‘liquidation date’ data (most rows were Nan) so we decided to just remove all companies not Active
  + Created a dictionary of all unique sector values to make sure all the values matched
  + Created a new data frame that is sorted by sector\_code\_1 and sector\_name\_1 called sector\_data
  + We then saved sector\_data as final\_cleaned.csv

***Industry\_cleanup*:**

* Contains: final\_cleaned\_data.csv
* Produces: industry\_added\_cleaned\_data.csv
* Steps:
  + Imported final\_cleaned\_data.csv as df
  + Within df[“sector\_code\_1”] we removed any spaces
  + We then imported all industry code files into the workbook
    - These industry files were scrapped from the UK government sector website and added to 20 individual csv files
    - They represent a grouping mechanism for all sector codes
  + After looking at all of the data on our visualization we decided to only include finance, sci & tech rsrch, and information tech related sector codes
  + After only including these relevant companies, into df we created a count column that would allow us to count each individual company
  + We decided to narrow our company pool down to specific sectors and called this data frame df1
  + We then noticed the sector names weren’t comprehendible and wouldn't show well on a visualization, so the names were replaced with simplified ones, without changing the meaning
  + After brainstorming on how to analyse the event data we decided to create a function called year\_month\_counts that would allow us to pick two different points in time and see the difference between sectors of company counts
  + We then saved this data as industry\_added\_cleaned\_data.csv

### Visualizations: *Mapping*

**tech\_city\_geospatial\_by\_sector**

* Contains: industry\_added\_clenaed\_data.csv, tech\_roundabout.csv
* Produces:
* Creates:
  + Geospatial chart (1)
  + Supporting Graph (1a)

**Impact\_events\_visualisation**

* Contains: industry\_added\_clenaed\_data.csv, tech\_roundabout.csv
* Produces: Code for pivot\_tech\_table\_cumcount.csv
* Creates:
  + Event data lollipop chart
  + Event heat map

**tech\_city\_growth\_over\_time**

* Contains: industry\_added\_clenaed\_data.csv, tech\_roundabout.csv
* Produces:
* Creates:
  + Timeline chart
  + Racing bar charts

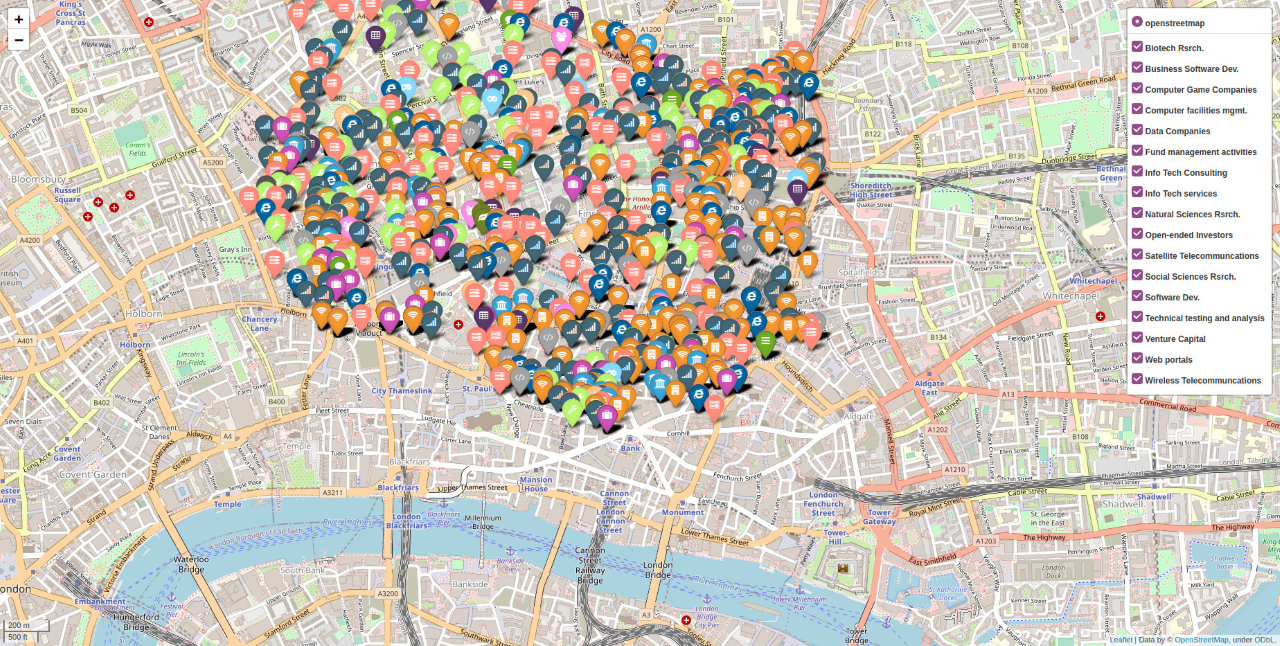
## 2.The Visualizations

**Overall Goal**: Visually depict the geographical composition, company growth, and impactful events affecting Tech City by sector. The visualizations are segmented into three categories that each explain an essential part of the tech city story.

### Geographical Visualizations

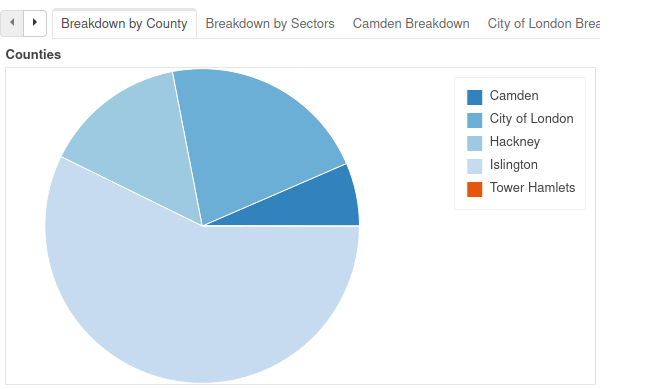
**Interactive Map**

* **Purpose:** To allow the viewer to observe the geographical composition of the sectors and companies that make up tech city.
* **Components:**
  + Hovertool: see each individual company name
  + Sector selections: view only selected sectors on map
  + Sector identification: sectors are color coded and represented by unique icons
  + Zoom feature: viewer can zoom in as far as he/she wants



**Interactive Pie Chart**

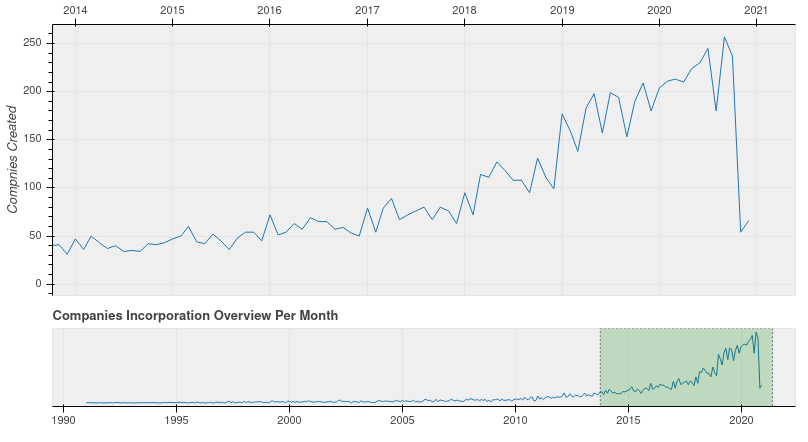
* **Purpose:** Supports the map by allowing the viewer to interact with and view sector and company counts for total district, total sectors, and more granularly by sector counts within each individual district of Tech City
* **Components:**
  + Interactive tabs: Shuffles through each categorization to see more specific data
  + Hovertool: Allows for identification of counts for each segment of the pie chart



### Timeseries Visualizations

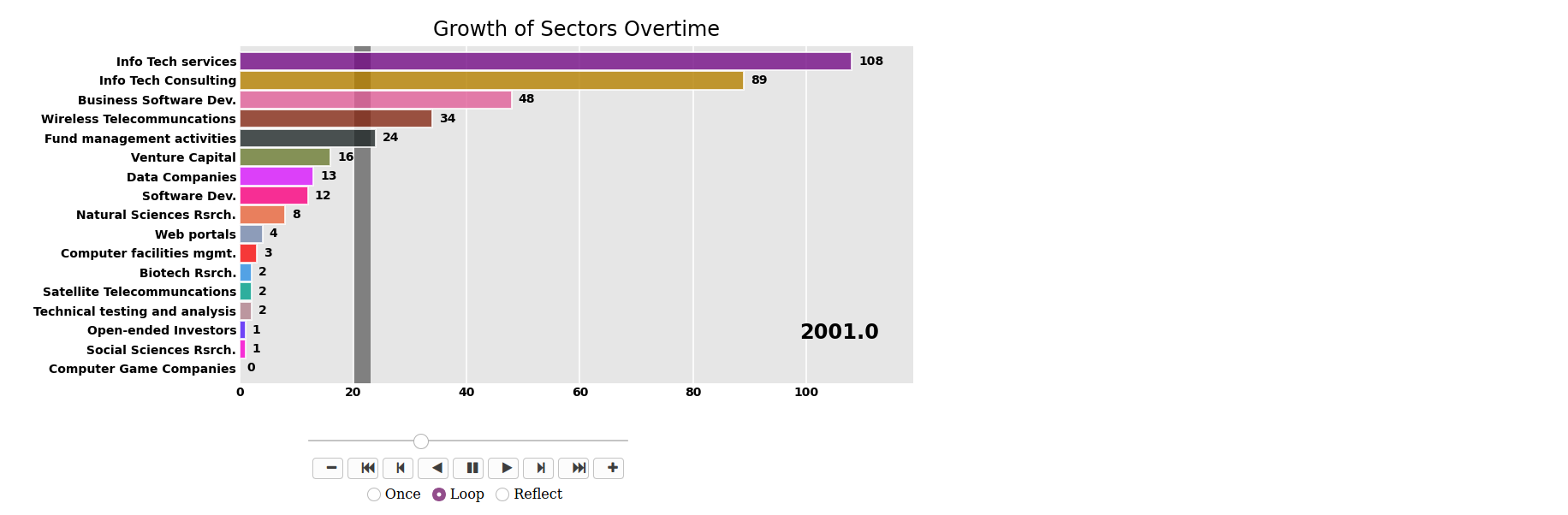
**Timeline Chart**

* **Purpose:** To show a simple overview of the total growth of Tech City over time. The viewer can zoom into specific points where they can see where growth was increasing or slowing.
* **Components:**
  + Timeline adjustment: To zoom into specific points in time



**Racing Bar Chart**

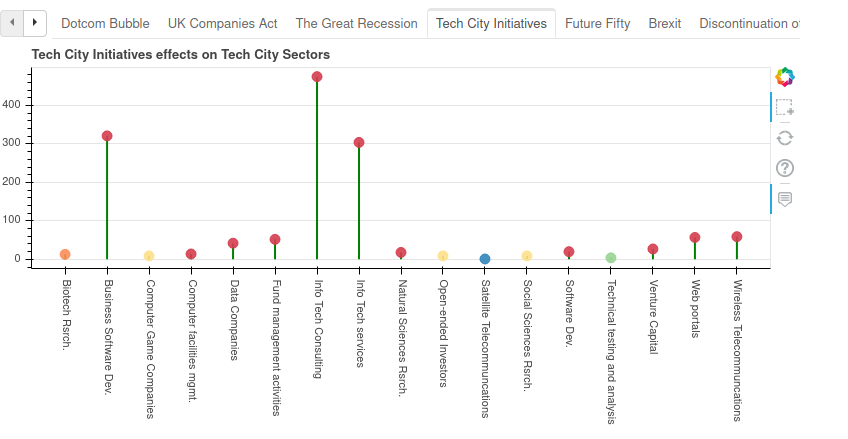
* **Purpose:** Provides a comprehensive understanding of the growth of all sectors in Tech City. The viewer can see sector growth from the inception of tech city to the most recent point in order to understand when companies were incorporated and during what dates in time.
* **Components:**
  + Can drag to any year after 1991 to see count of companies in each sector of Tech City
  + The grey bar represents the mean so we can see which industries are outperforming and underperforming in terms of growth.



### Impactful Event Visualizations

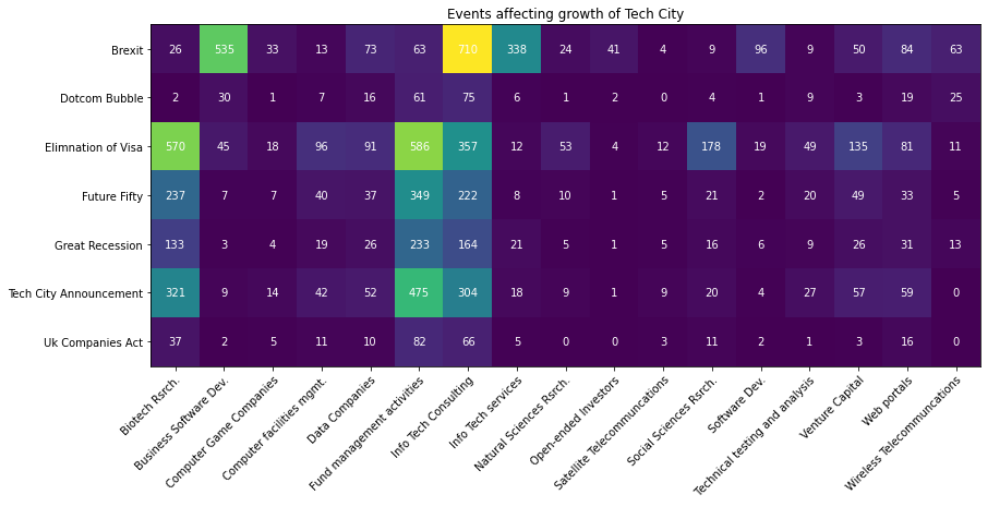
**Event Lollipop Chart**

* **Purpose:** Shows a breakdown of each impactful event’s effect on company growth by sector. The viewer can identify which events helped or hurt sectors based on how much that sector grew. The viewer can consider the total count of companies incorporated during this time period or the percent growth during this time period.
* **Components:** 
  + Hovertool: allows the user to hover over the dot and see percentage growth
  + Tabs that represent each major event



**Event Heatmap**

* **Purpose:** Simple visualization that provides the viewer with color coded panels of events that cause company incorporation rates to spike.



## 3.Impactful Event Overview

### Industries with the Most Impact:

* **Biotech Research**: This industry involves research into biological processes, cells, organisms and more to generate new technology that can aid in areas such as agriculture.
* **Fund Management Activities:** This industry involves the allocation of assets from individuals or organisations to invest in other projects or firms and potentially grow the fund.
* **Info Tech Consulting:** This industry focuses on providing advisory services to firms on how to efficiently incorporate IT solutions to meet business objectives

### Events:

* **Dotcom Bubble**: This encouraged only 2 Biotech companies to open in the area, meaning it did not disrupt the research conducted. With the inception of the internet, investors were keen to invest and capitalise on this new technological transformation and, hence, Fund Management firms reached 61. The fast disruption of business models using the internet instigated a demand for advisory on technological solutions to aid firms, leading to the IT Consulting industry with 75 firms in the area.
* **UK Companies Act:** This was introduced as the first compressive set of rules to provide a formal structure to how companies would conduct operations in different sectors today. These regulations acted as a guide for firms and motivated the Biotech and Fund Management industry to expand, with 37 and 82 (increase of 1750% and 34%) companies respectively in the area. However, IT consulting firms had reduced to 66 (12%) due to the act established not being able to properly regulate new technology startups’ infrastructure and operations.
* **Great Recession:** Biotech companies grew to 133 in the area. This is due to the nature of the industry affecting areas such as agriculture, which in turn is a necessity for survival. Sectors revolving around food tend to perform well during recessions and therefore, a growth of 259% in companies occurred. The uncertainty of this period required businesses to seek outside help on technological solutions that would be cost effective and maintain survival/growth. Therefore, IT consulting firms grew to 164 (148%). While the economic downturn slowed down the economy, it provided a large opportunity for investors to buy. Therefore, firms in this sector grew to 233 (184%).
* **Tech City Announcement:** The announcement involved a lot of promotion, including the TweetDeck, a dashboard launched by Twitter that boosted the reputation of Tech City. The publicity encouraged more successful startups to establish themselves here. Therefore, there is a significant growth in the Biotech, Fund Activities and IT Consulting industries (141%, 103%, 85%).
* **Future Fifty:** This was an initiative launched by the government and Tech City to support late-stage technology firms in building a collaborative and supportive network. The aim of the programme was to accelerate growth of digital companies in their growth stage. However, the Fund Activities, Biotech and IT Consulting industry had seen a reduction of 26%, 26% and 27% respectively. Despite the goals of this initiative, the programme had backfired in these top 3 industries and in the others as well.
* **Brexit:** The uncertainty and crash of the stock market resulted in a large decrease in a similar drop in companies of 2 sectors (89%- Biotech, 82%-Fund Management). However, the IT Consulting industry saw the opposite, with a significant increase in companies from 222 to 710 (220%). Similar to the recession, the uncertainty of Brexit impacts and continuous advancements in technology required businesses to gain outside help and IT solutions from technology advisory service firms.
* **Elimination of visa:** In March 2019, the UK innovator visa replaced the entrepreneur visa to target more experienced and established business individuals. Moreover, the parameters to gain the visa was reduced; requirement to invest £200,000 had been changed to £50,000. Subsequently, 2 industries experienced a large increase in the number of companies joining Tech City (2092% Biotech and 830% Fund Activities). However, IT Consulting startups were not able to benefit from this and saw a drop of 50%. This is due to heavy competition from established business models in tech advisory firms such as PwC and Deloitte, who also established strong relationships with customers.