

5. Dynamic and Interactive Visualisation using Tableau

In this section we turn our attention to the use of **dynamic interactive graphics** using Tableau. Dynamic graphics can be considered ‘work in progress’ illustrations and are valuable aids in exploring data sets. The objective of exploratory analysis as outlined in an earlier section is to discover interesting patterns and structures in data sets that perhaps would remain hidden using more conventional approaches to data analysis. This approach to data analysis has been made possible in recent years through the development of powerful interactive software like DataDesk and Tableau. In a sense exploratory data analysis can involve the creation of many graphics in the exploratory process. These graphics are viewed by one analyst in contrast to one graphic viewed by many, say in a publication, which is the case for static graphics. In this section we will examine some dynamic and interactive features of Tableau using the electoral division data set examined earlier.

5.1 Worked Example

The electoral division data contains all of the **Theme 10 education variables** from the 2016 CSO census file. In this example we investigate the interactive features of Tableau using table 4 of the Theme 10 education series which lists **population aged 15 years and over by sex and highest level of education completed**. The particular variable we will use is **T10_4_ACCAT** which is the **number of persons per electoral division who have advanced certificate/completed apprenticeship**. This data is provided in the Excel sheet **ED2016.xls** in the **Theme 10 Extract** folder in **Mapping Files**.

Figure 5.1 shows the rate per 10,000 population of ACCAT in each electoral division. The plot was obtained by joining the ED shape file **Electoral_Division_CSO_Generalised_20M** with the CSO census file **CensusED2016**. The rate was obtained using a calculated field by dividing **T10_4_ACCAT** by **T1_1_Agett**, multiplying by 10,000 and encoding using the colour mark.

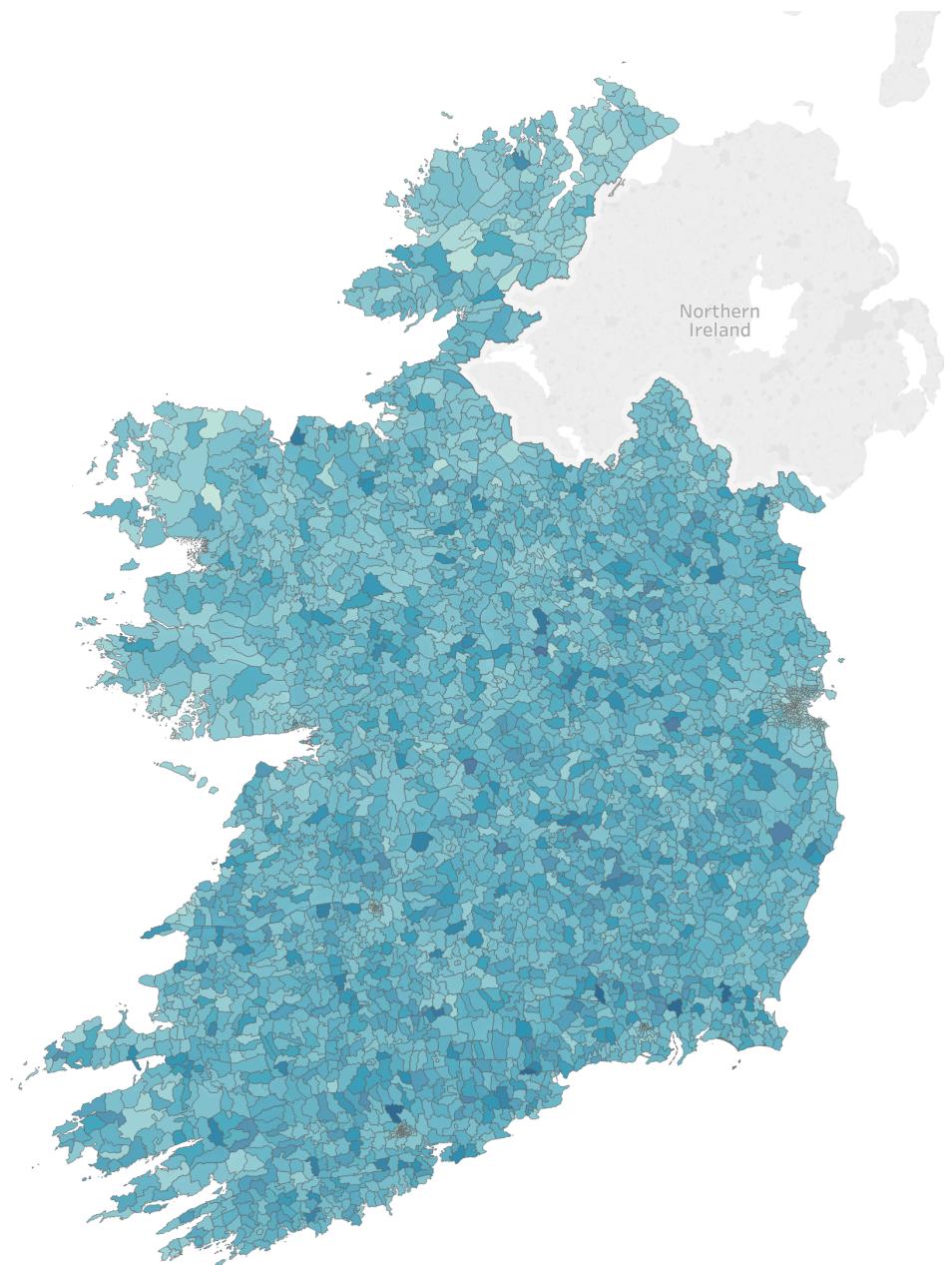


Figure 5.1: Rate per 10,000 of ED population who have advanced certificate/completed apprentice

In the next section some of the dynamic and interactive features of Tableau will be explored.

Exercise

Using the shape and Excel files create the plot shown in Figure 5.1 of the rate per 10,000 population in each electoral division for [**T10_4_ACCAT**](#).

5.2 Dynamic Interactive Graphical Tools

1. Tooltip

To add the name of the Electoral Division to the map shown in Figure 5.1 we can place **Ed English** onto the **tooltip** icon in the **Marks** dialog. As we mouse over the map the name of the electoral division appears on the tooltip together with the values of the other variables used to create the plot as shown in Figure 5.2.

Visualising data interactively allows for fast regional comparisons to be made. The ability of make fast comparisons is a fundamental component of dynamic and interactive analysis and can lead to the discovery of interesting patterns - in the words of Edward Tufte we must always keep in our mind the phrase *compared to what?*

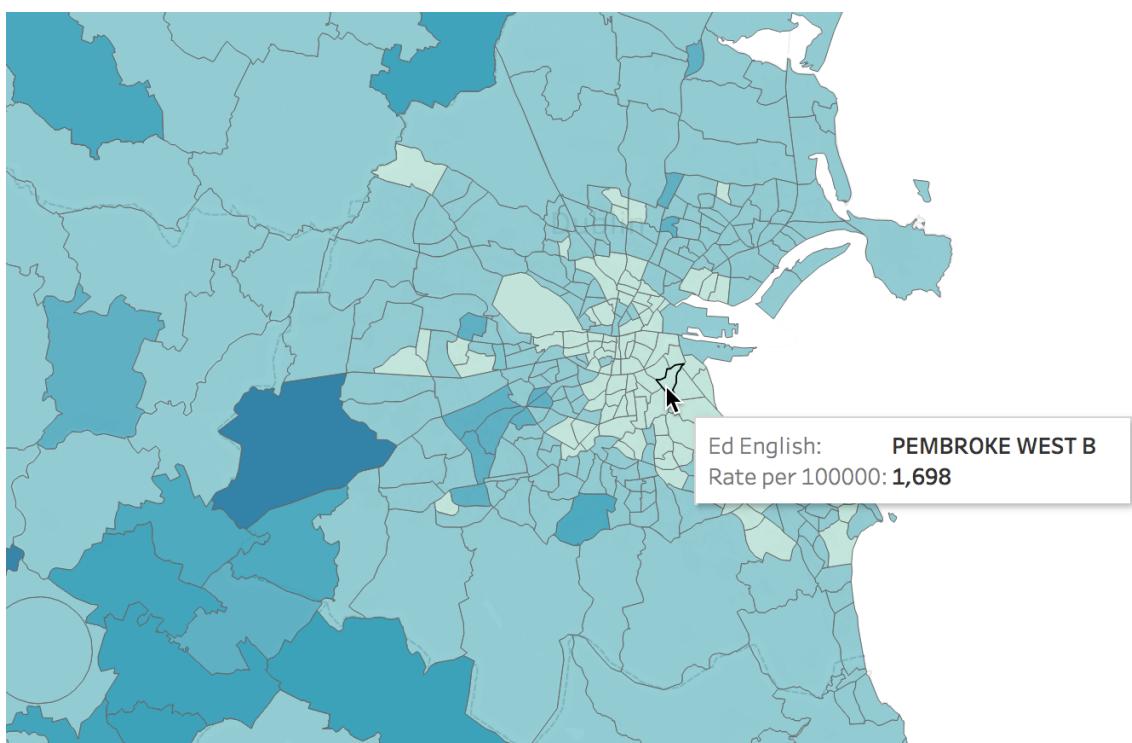


Figure 5.2: Using the tooltip for dynamic interaction

2. Creating Filters

Filters are a great way of analysing subgroups interactively and dynamically. They can be usefully applied to discrete variables where each tick box or radio button representing the values in the filtered variable can be selected and displayed.

For the Theme 10 data set we can create a filter for county by selecting the variable **countyname** variable and then the option **Show Filter** in the drop down menu at located at the end of the pill. The filter now appears to the right of the graph with different possible options including tick boxes and radio button displays. Selecting **single value list** provides a panel of radio buttons which when selected show the distribution for the selected county as shown in Figure 5.3.

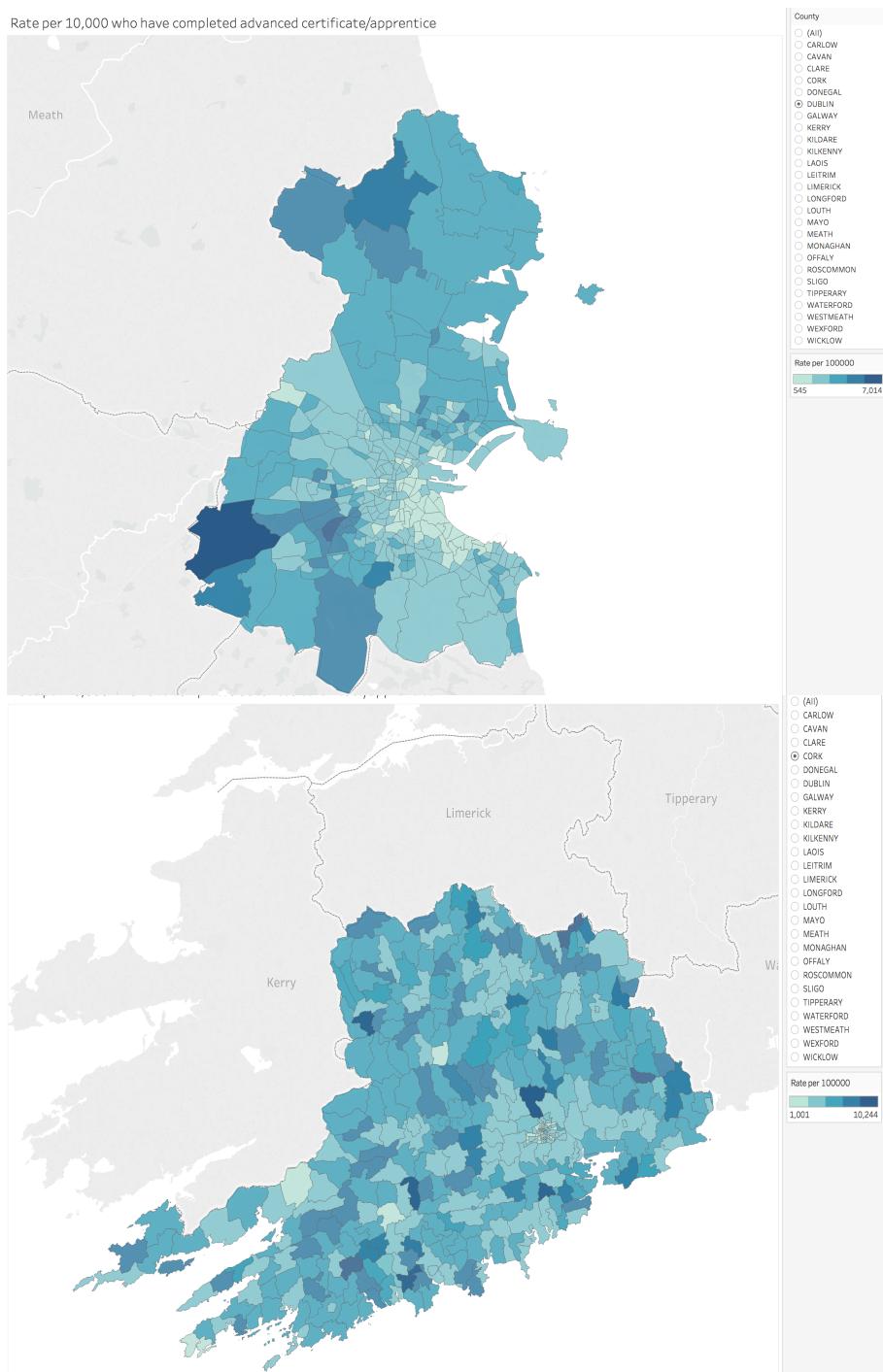


Figure 5.3: Selecting Dublin (top) and Cork (bottom)

Deselecting the **fix map** icon on the menu bar will show just that county on the display. If the analyst wants to display a number of counties use the **multiple values list** filter option. The top plot in Figure 5.3 is the distribution of the rate (colour encoded) of ED population who have completed an advanced certificate or completed an apprenticeship. The bottom plot is for Cork County.

3. Using the selection tools to create subgroups

It is possible to directly select cases using the selection tools of Tableau, create a subgroup which can in turn be filtered to allow dynamic interaction. For example, suppose we want to investigate the distribution of T10_4_ACCAT in two geographic regions that are not coded in a variable - say Dublin north-city and Cork south-city. Firstly, select one of the selection tools e.g. the lasso tool and click and drag on the Dublin north region. When the area is selected a pop-up window appears the top of which contains a **paperclip** icon. Selecting the paperclip puts the selected points into a variable called **ED English (group) 1** with a name consisting of the first few electoral divisions and displayed in a legend located to the right of the graph as shown in Figure 5.4.

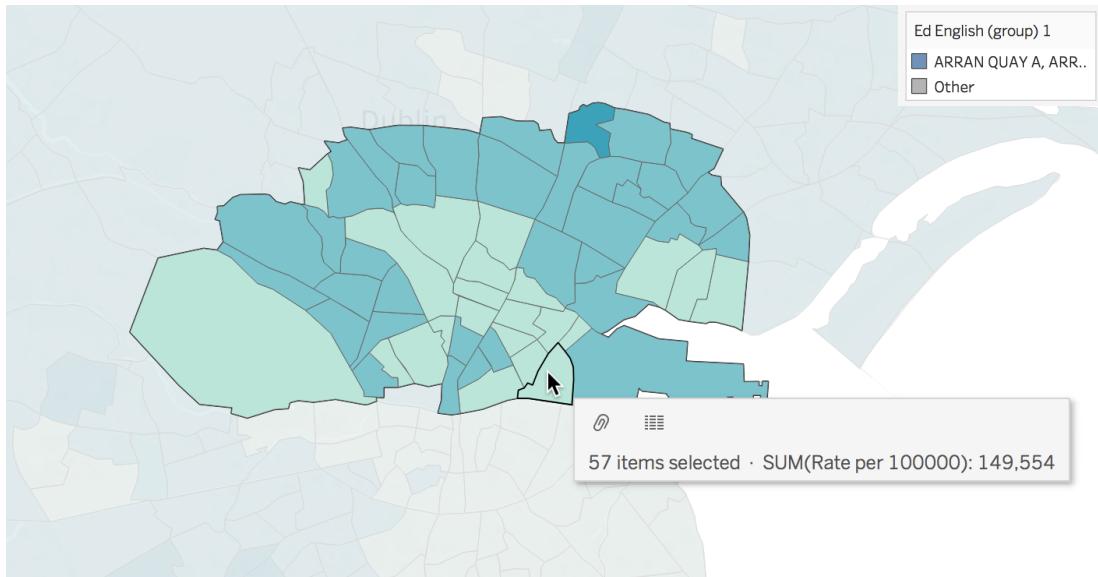


Figure 5.4: Selecting Dublin north-city using the Lasso

Note: Please ensure that G UID is placed on the **level of detail field** otherwise the paperclip icon will not appear.

Next select the cork south city region and repeat this process once again. Selecting the paperclip icon adds a new list of values to the **ED English (group) 1** as shown in Figure 5.5.

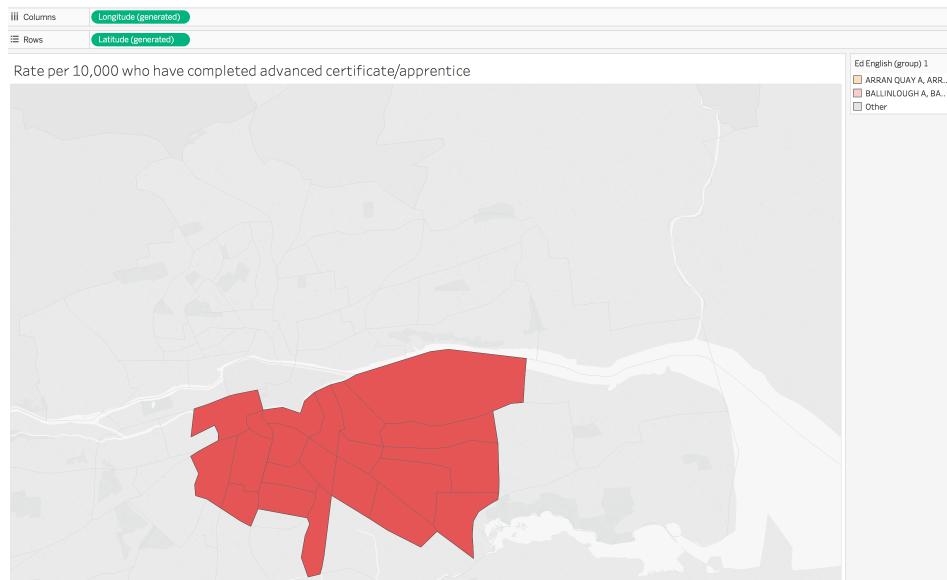


Figure 5.5: Selecting Cork south-city using the Lasso

The ED English (group) 1 can be found in the dimensions panel and clicking on the drop down menu at the end of the pill we can select **edit group**. A menu appears that allows the user to change the group names into more meaningful labels as shown below by selecting the group and then the **Rename** button. Note that Tableau puts the rest of the towns not selected into an **other** category which has been changed to **Rest of Ireland**.

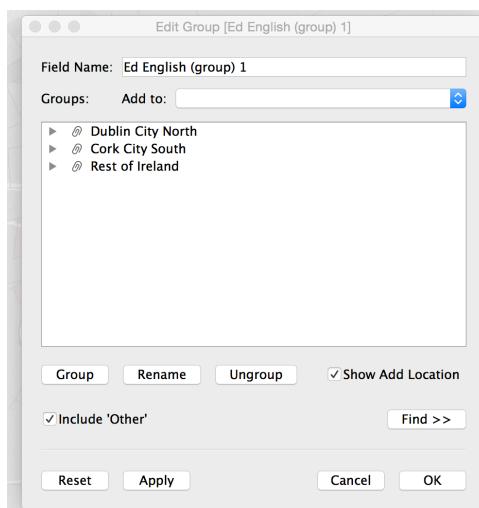


Figure 5.6: Editing the value names of the grouping variable

The name of the grouping variable ED English (group) 1 can be also edited by selecting the variable and from the drop down menu and select **edit title**. In this example we have renamed it to the more descriptive **Regional Analysis**. Now that the groups have been set up we can interact with the displays and assess differences that may exist between the regions.

We can do this interaction in a number of ways. The first way is to turn the **highlighter** on. This is located on the top the legend containing the group names located to the right of the map. In the dropdown menu select **Highlight Selected Items**. Selecting each group now shows the corresponding regions.

Another way is to create a filter by selecting the grouping variable Regional Analysis in the dimensions panel and then **show filter**. This allows a larger number of options for interacting with the groups. Selecting **single value list** provides a radio button for each area which when selected will zoom to the area as shown in Figure 5.7.

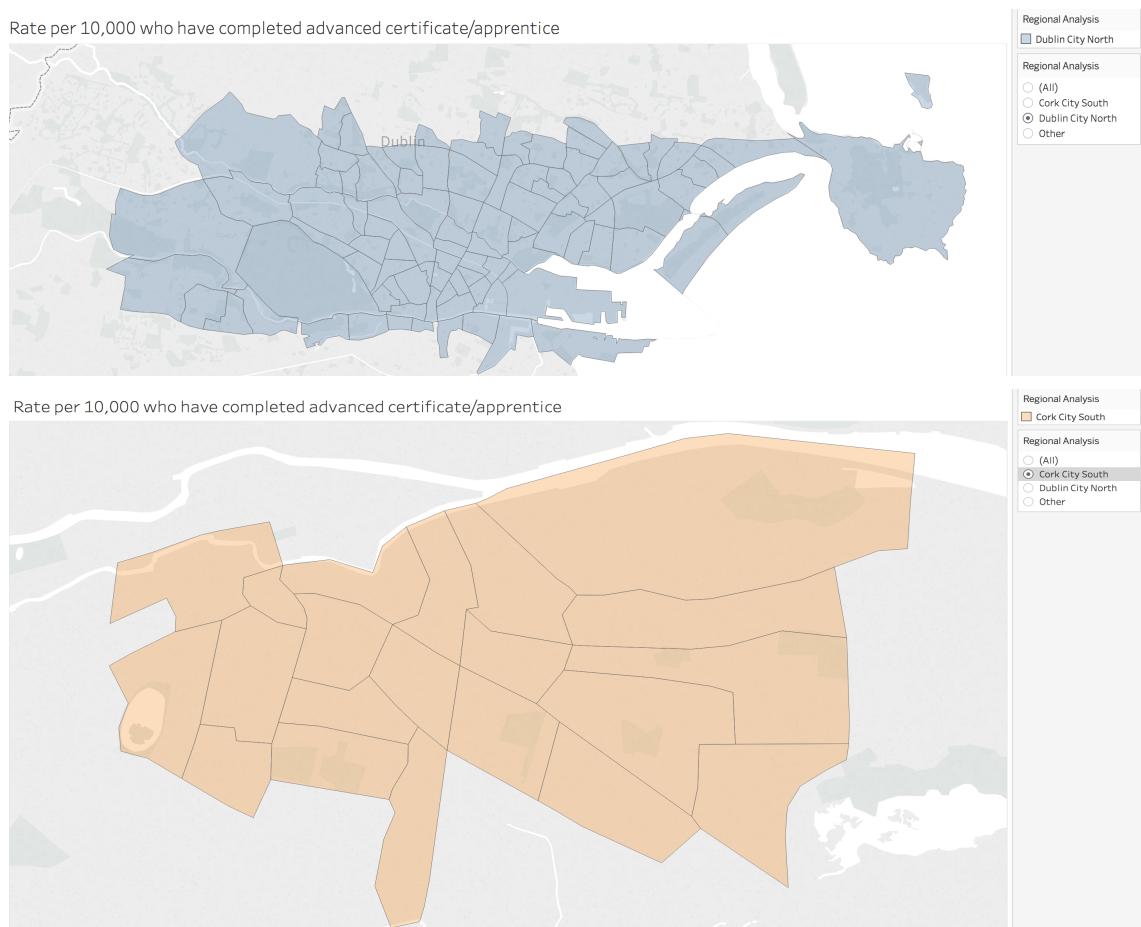


Figure 5.7: Selecting Dublin north city and Cork south city using the filter

This filter can now be applied to graphs created in other worksheets. For example, if we create a box plot of the rate of T10_4_ACCAT and then create a filter from **Regional Analysis** we can then visualise the rate who have obtained a higher certificate/completed apprentice qualification for each of the three regions as shown in Figure 5.8:

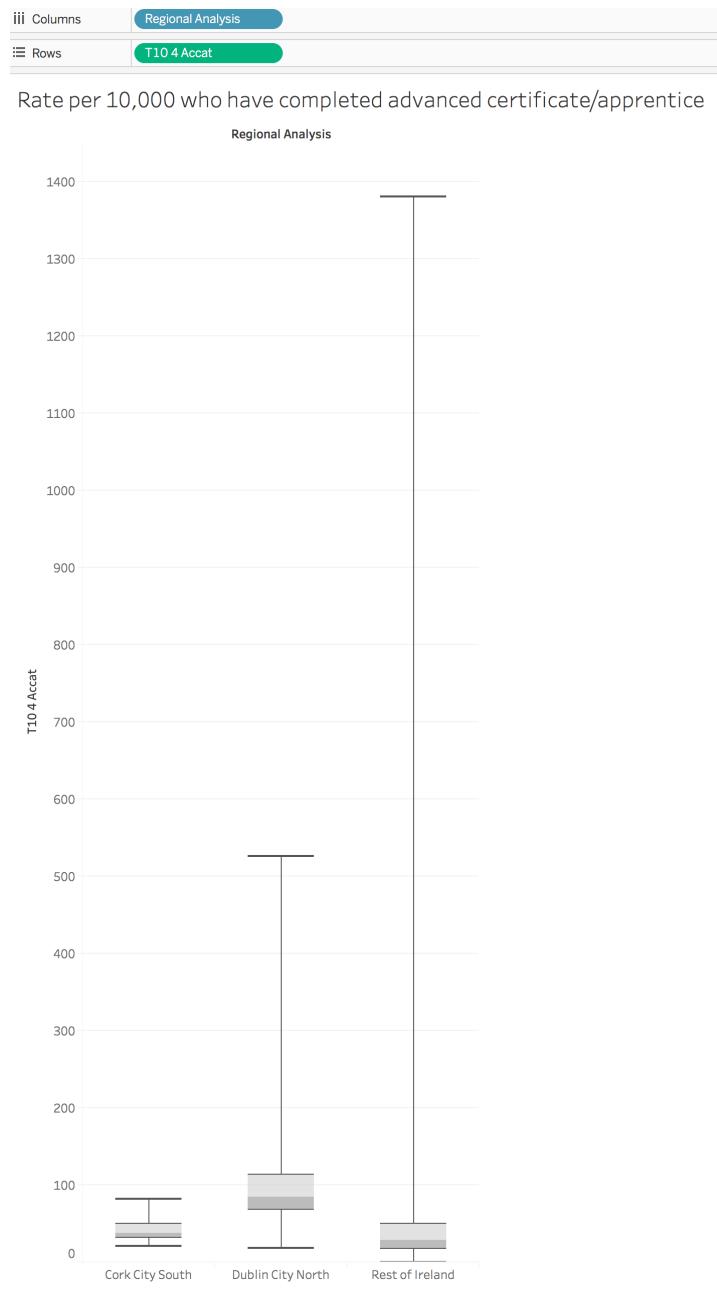


Figure 5.8: Applying the filter to other worksheets

The data underlying the above graphics can be easily exported from the menu item **worksheet**, selecting **export** and then **data**.

5.3. Creating Dashboards in Tableau

Dashboards are windows that contain a number of graphs that can be linked to allow for dynamic interaction. They are created by selecting the **Dashboard** tab in Tableau or selecting **New Dashboard** from the Dashboard menu. Icons representing content that has been created in worksheets will now be displayed on the left hand side of the blank dashboard window. These icons can be selected and moved to the canvas allowing graphs and tables to be displayed. We will now illustrate the creation of a dashboard using the graphics created in the previous section based on the number per small area who have an advanced certificate or completed an apprenticeship.

Open the saved Tableau file which was used to explore the variable T10_4_ACCAT earlier. Selecting the **dashboard tab** we see two worksheet icons on the left of the screen. Each worksheet contains a graph - one a map of Ireland with filters for all the counties of Ireland as shown in Figure 5.9. The other worksheet is a box plot of the rate per 100 of ACCAT for each county as shown in Figure 5.10.

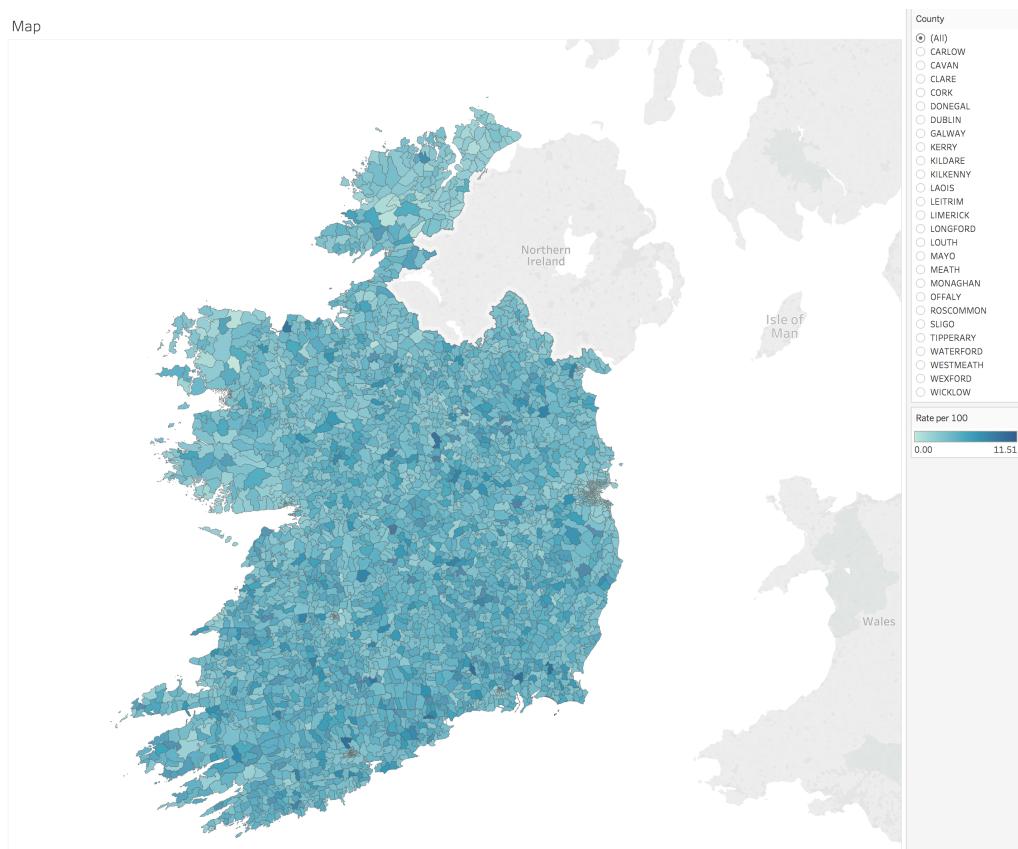


Figure 5.9:Rate per 100 of ED population who have ACCAT

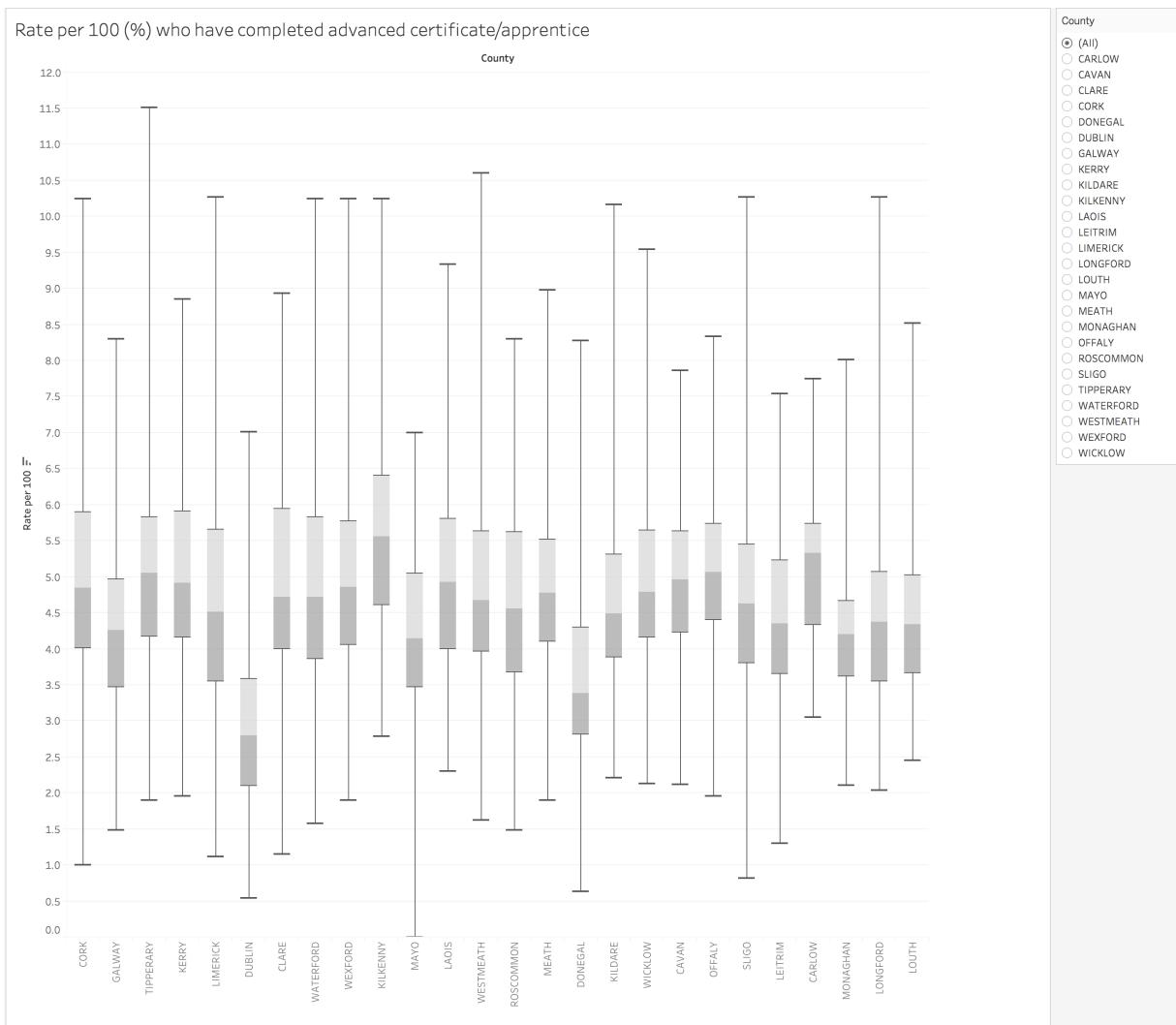


Figure 5.10: Rate per 100 of county who have ACCAT

We can now drag each worksheet onto the dashboard window and obtain the dashboard shown in Figure 5.11 where each county/city can be filtered to show the geospatial distribution of ACCAT by electoral division in each county/city together with an overall visual summary of the rate per county/city using a box plot representation.

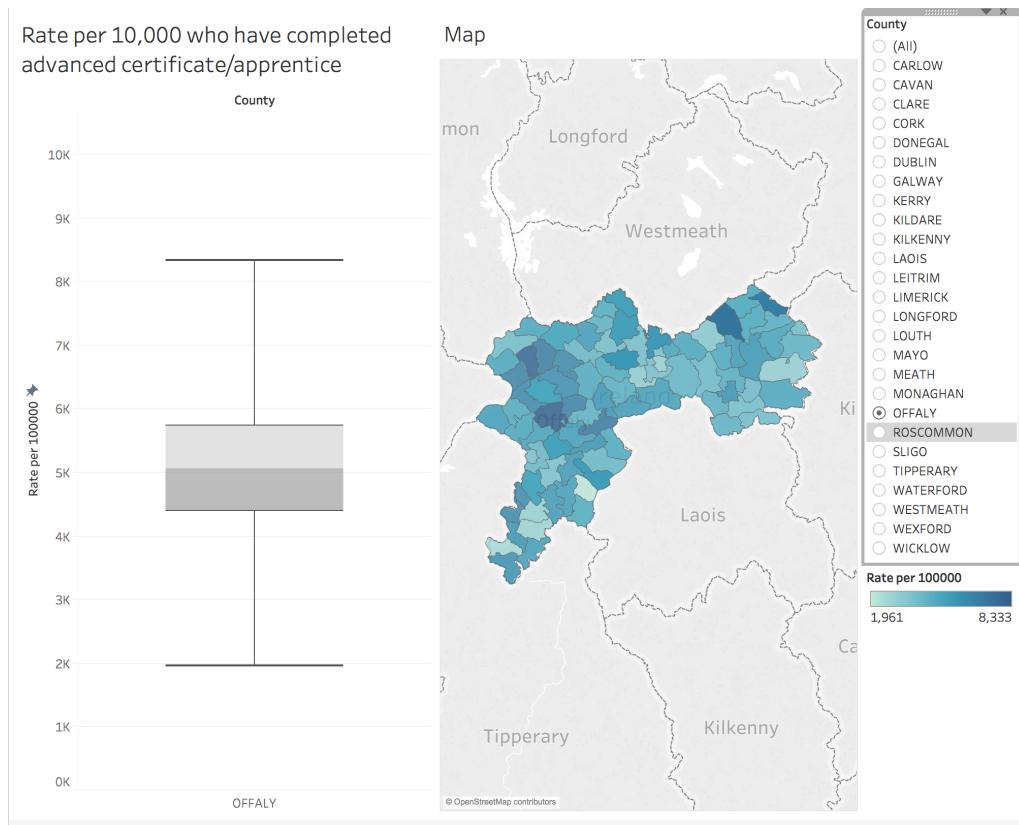


Figure 5.11 Dashboard of ACCAT

If we select the dropdown menu item from each of the filters and then **apply to all plots** this allows each filter to operate on both plots simultaneously. For example the dashboard in Figure 6.1 shows the data for Offaly (selected from the county filter) in the **Rest of Ireland** category (selected from the Tourist Regions Filter). Now that we have our dashboard we can share it in a number of ways with other users and organisations.

Exercises

1. Theme 13 in the CSO Census 2016 examines the occupations of citizens by small area. The variable **T13_1_STOT** records the number of persons employed in **Skilled Trade Occupations** which is located in the excel file **CensusSA2016.xls**.
 - a) i) Visualise the geographical distribution of this variable by Small Area (SA) by joining the shape file **Small_Areas__CSO_Generalised_20M** with the CSO Census file **CensusSA2016.xls**.
 - ii) Create a calculated field of **T13_1_STOT** per 100 SA population and use encode the map created in i) by colour using this calculated field.
 - iii) Design some interactive features into your map as follows:
 - Allow users to visualise the name of the small area as they mouse around the map.
 - Create a county filter that zooms to the selected county when selected by the user.
 - Using the selection tools create a group that includes three regions corresponding approximately to north-west, south-east and south-west of Ireland. Label the group **region** and label the filter values **north-west**, **south east**, **south west** and **rest of Ireland**.
 - Create a filter of **regions** that zooms to each chosen region on selection.
- b) i) Create a dashboard using the worksheets created in a ii) and a iii). This dashboard should contain a dynamic filter for **region** for both plots.
- ii) Create a trellis plot in a new worksheet where each panel represents a box plot of **T13_1_STOT** per 100 SA population for each region
- iii) Write a note summarising any observed differences in the distribution of **T13_1_STOT** per 100 SA population for each region.

2. The data in the Excel worksheet **ecoli4D** in Data(2016) records location, gender and diagnosis of three diseases Cryptosporidiosis, Verotoxigenic Escherichia coli infection and Giardiasis during 2015 and 2016 in the Greater Dublin Area. The location data has been jittered by adding noise to the longitude and latitude of each location in the form of a standard normal probability distribution. Please note that the data set is for illustrative purposes only as it excludes some locations for which the longitude and latitude were not determined.
- a) Create three worksheets in Tableau for this data set that visualise:
 - i) geographical distribution of ecoli in the Greater Dublin Area
 - ii) Incidence of ecoli over time
 - iii) gender and age distribution associated with ecoli
 - b) Create a dashboard using the worksheets created in a). This dashboard should contain a dynamic filter for ecoli disease category.
 - c) Using the results of a) and b) write a short note that summarises the incidence of ecoli paying particular attention to any observed contrasts in incidence in time, location gender and age.

Please save your Tableau file as a .tbdx file.

- 3. a)** The data in the Excel worksheet **RoadAccident** in Data(2018) contains approx 120,000 accident records. Each record provides the longitude, latitude, county, year, type of accident and primary collision type for reported accidents that occurred between 1996 and 2013.

Create four worksheets in Tableau for this data set that visualise:

- i) geographical distribution accidents in Ireland the Greater Dublin Area
 - ii) distribution of the primary collision type
 - iii) distribution of accidents by weekday
 - iv) yearly road accidents by Type over the years 1996-2013
collision
- b) Create a dashboard using the worksheets created in a). This dashboard should contain a dynamic filter for county that dynamically visualising the distribution for the selected county in all dashboard plots.
- c) Explore the dashboard by navigating through the county filter and write a short note that summarises road accident incidence highlighting any national or county features you think are interesting.

4. The data in the Excel worksheet **tourism** in Data(2018) contains the variables provided in the table below:

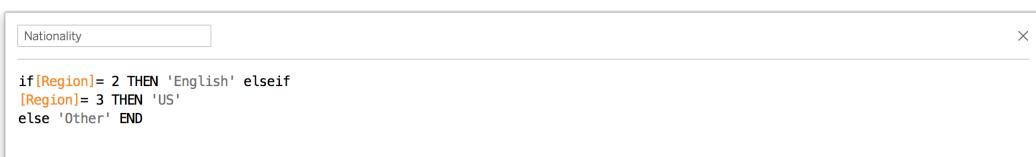
Variable Name	Description
Holiday trips	Number of holiday trips
Business trips	Number of business trips
VFR trips	Number of trips involving V isits to F riends and R elatives
Town	Name of Town Visited
Longitude	Column Value (x-axis)
Latitude	Row Value (y-axis)
Country of Residence	Britain (value 2), North American (value 3) and Mainland Europe (value 4)
Brand*	Wild Atlantic Way, Ireland's Ancient East and Dublin - A Breath of Fresh Air

The research department of Failte Ireland are interested in determining if the geographic distribution of visitors to Ireland varies according to country of residence. They would also like to obtain further insight on how visitors to the tourism brands (Wild Atlantic Way, Ireland's Ancient East and Dublin - A Breath of Fresh Air) vary by segment (business, holiday and VFR) and by country of residence (Britain, North American and Mainland Europe).

This exercise will feed into Failte Irelands strategic plans for growing the business in 2017 onwards through identification of potential opportunities for growth.

- i) State giving a reason why if this data set is 1D, 2D or MD?
- ii) List each of the variables in the data set and state if they are discrete or continuous?
- iii) Import the data from Excel into Tableau. Create a calculated field that recodes the numbers 2 to 4 in the variable *Country of Residence* to text labels *Britain*, *North American* and *Mainland Europe*. Call this new variable **Residence**.
- iv) Using Tableau create another variable called **Brand** that contains three values *Wild Atlantic Way*, *Dublin - A Breath of Fresh Air* and *Irelands Ancient East* (which is the remainder of the country). The values can be selected approximately using the selection tools. Label each of the selected regions as above and call the variable containing the groups **Brand**.
- v) Create three worksheets - one for each of holiday, business and VFR that map the geographical distribution of towns visited. Incorporate dynamic interaction to each worksheet by using **Residence** as a quick filter.
- vi) Create a fourth worksheet which visualises the distribution of segment (business, holiday and VFR) by brand (*Wild Atlantic Way*, *Dublin - A Breath of Fresh Air* and *Ireland's Ancient East*) and by residence (Britain, Mainland Europe and North America). This visualisation need not necessarily be map based.
- vii) Create a dashboard that contains the plots created in v) and vi).

Note: To recode values in variables to more meaningful terms we can use calculated fields which were introduced earlier in the previous sections. In this example the variable **region** has codes 2, 3, 4 and 5 which refer to tourists from England, North America, Mainland Europe and Other Regions. To convert say region codes 2 and 3 to English and US we select **Create Calculated Field** from the **Analysis** menu. In the dialog box that appears enter the following logic commands:



```
Nationality
if [Region]= 2 THEN 'English' elseif
[Region]= 3 THEN 'US'
else 'Other' END
```