



Choosing a Data Visualization Tool

Which Tools

General Data Presentation Tools

Excel	Excellent for tables and many graphs (.gov)
PowerPoint	Excellent to produce a story about data.

Specialist Data Analytics Tools

R	Precise graphics for specialists.
Python	Needs coding knowledge to be effective.
Knime	Open Source drag and drop analytics.
Alteryx	Widely used drag and drop analytics.
Tableau	Widely used, better analytics and mapping.
PowerBI	Widely used, easier interface, better layout.

*Tool costs and license arrangements frequently change.

General Purpose Tools for Data Visualization

PowerPoint

Widely used presentation software.

Has no knowledge of underlying data, purely a drawing package.

Good for creating linear stories about data.

Can combine and annotate outputs from any other tool.



Excel



Department	Black	Gold	White	Total
Phones	244,099	60,000	50,000	304,099
Tablets	100,000	100,000	89,340	200,000
Laptops	30,083	10000	10000	40,083
Totals	£ 374,182.00	£ 170,000.00	£ 149,340.00	£ 544,182.00

Highest sales

Lowest sales

Spreadsheet tool that has some knowledge of the data and analytics capability.

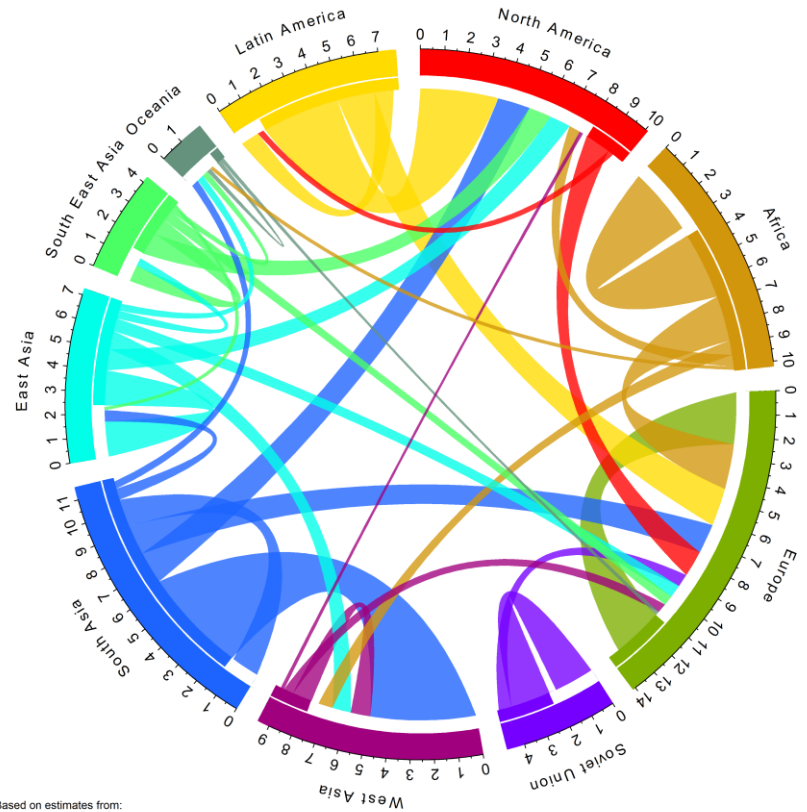
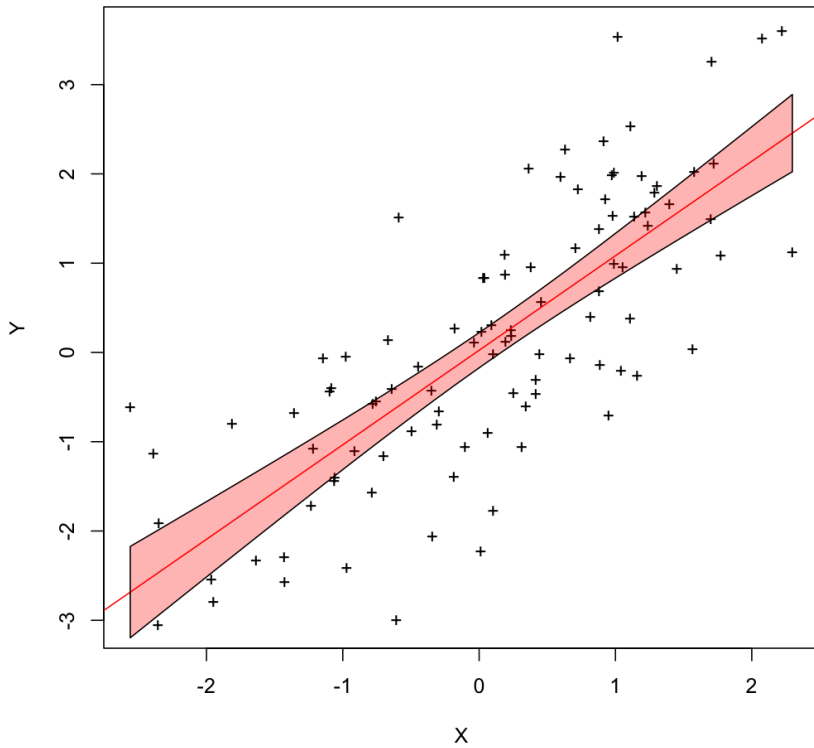
Effective table formatting tools.

Range of graphing and charting tools plus external plugins.

Specialist tools for Data Visualization

R

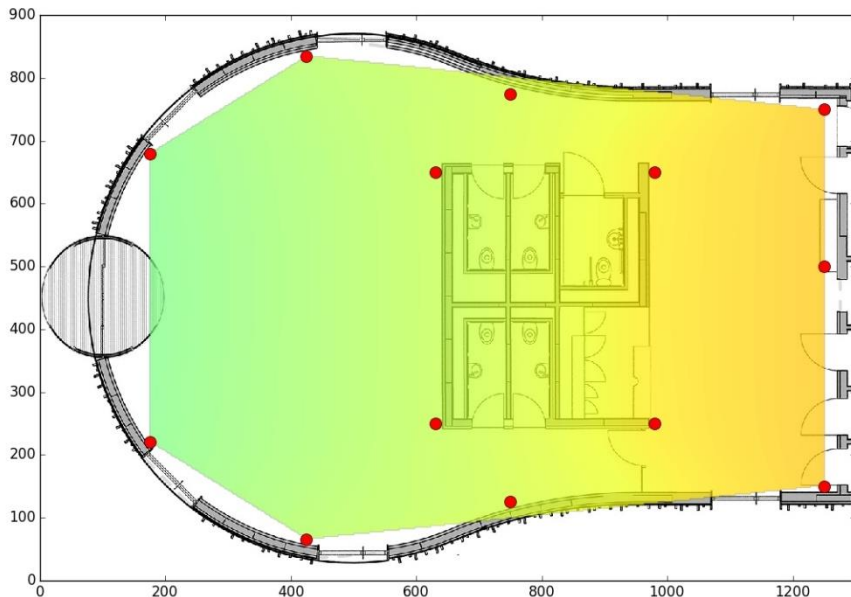
R provides a package of tools for data analytics which can provide precise control over graphics, particularly if you are happy to code.



Based on estimates from:
Abel and Sander (2014) *Science* Vol. 343 no. 6178 pp. 1520 – 1522

Python

Programming language that glues together a wide range of libraries widely used for data analytics and visualization.



```
#Generate a resampling grid of regular points to create the heatmap in
grid_x, grid_y = mgrid[0:1300:520j, 0:900:360j]

#Read in the irregularly placed sensor data from a file
print "CWD is: " + os.getcwd()
print "Starting file read"

tempData = csv.reader(open('tempData.csv', 'rU'), delimiter=',')
tempData.next()
x = []
y = []
t = []

#Loop through the data
for line in tempData:
    x.append(line[0])
    y.append(line[1])
    t.append(line[2])

#Convert from the sample points (x,y) to an interpolated heat map on a regular grid
(grid_x, grid_y)
print "Creating heatmap using griddata"
resampled_temp = griddata( (x,y), t, (grid_x, grid_y), method = 'cubic')

#Create the visualization
print "Creating visualization"

img = imread( "theKey_Half.png" )
plt.subplot(1,1,1)

# Create labelled axes
plt.axis([0,1300,0,900], zorder=2)

#Plot the sensor locations as red circles.
plt.plot( x, y, 'o', c="red", ms=10, zorder=2)

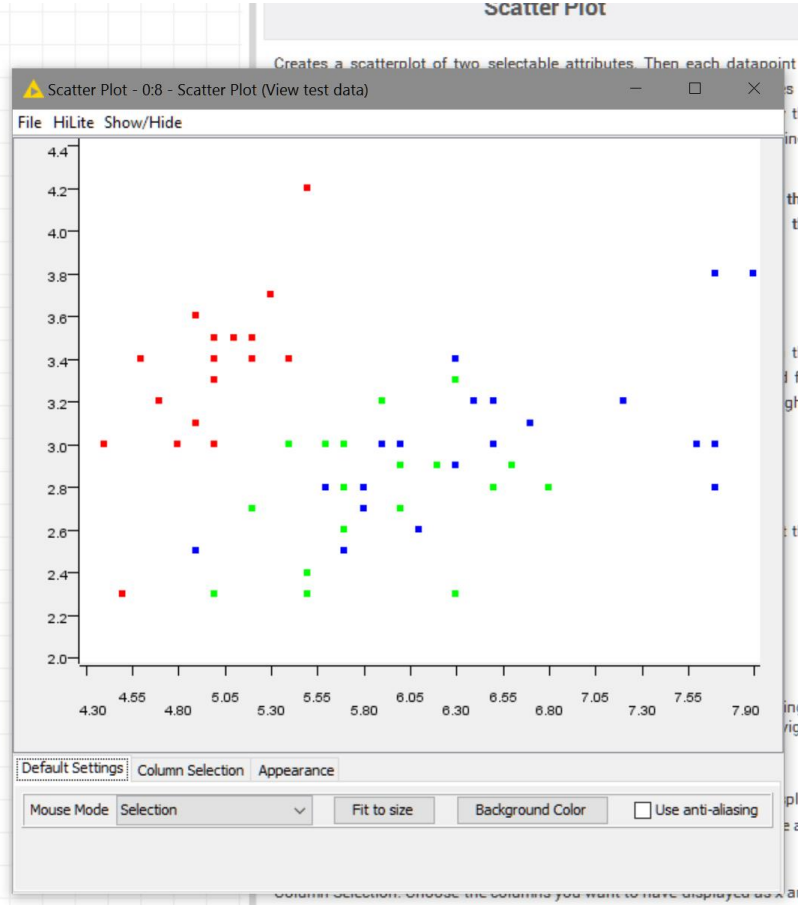
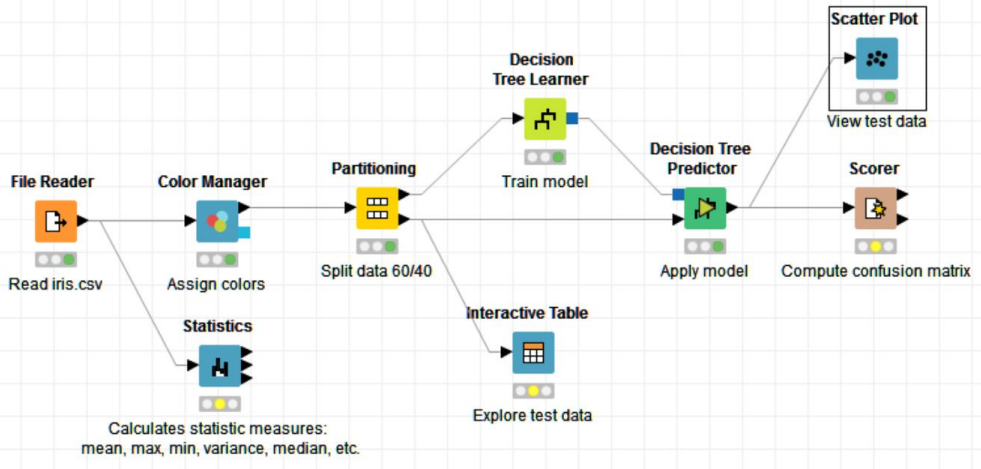
#Plot the key building background image (at z=0 so it is behind everything else)
plt.imshow(img, cmap="gray", interpolation="none", zorder=0)

#Plot the resampled temperature data.
# vmin and vmax set the minimum to maximum temp range for the default color map.
plt.imshow( resampled_temp.T, vmin=10., cmap=colormaps.viridis, vmax=25.,
            extent=(0,1300,0,900), origin='lower', alpha=0.7, zorder = 1)

#Show everything in blocking GUI window
plt.show()
```

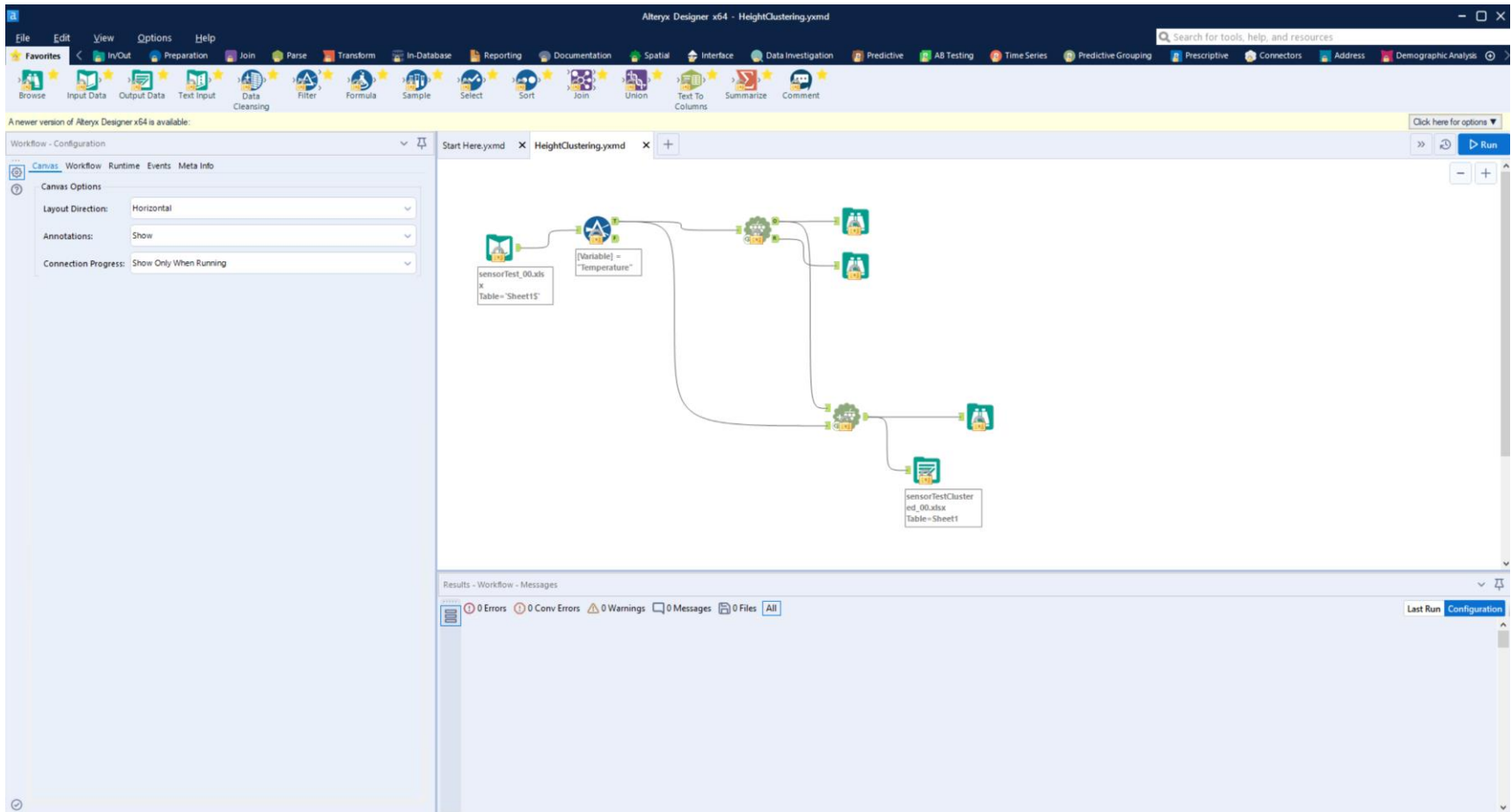

KNIME

This Example Workflow uses a **File Reader** node to import the Iris dataset (included). It then assigns visual properties with a **Color Manager** node and computes some basic statistics with a **Statistics** node. The data is split into training and testing fractions with a **Partitioning** node. The **Decision Tree Learner** generates a predictive model in PMML from the training fraction which is then applied to the test fraction using the **Decision Tree Predictor**. Model performance is evaluated with a **Scorer** node, which is applied after the **Decision Tree Predictor**. Finally, errors can be explored interactively, by using an **Interactive Table** node to highlight certain classes of errors which can then be visualized using a **Scatter Plot** node.



Open source well supported drag and drop toolset for data analytics.

Alteryx



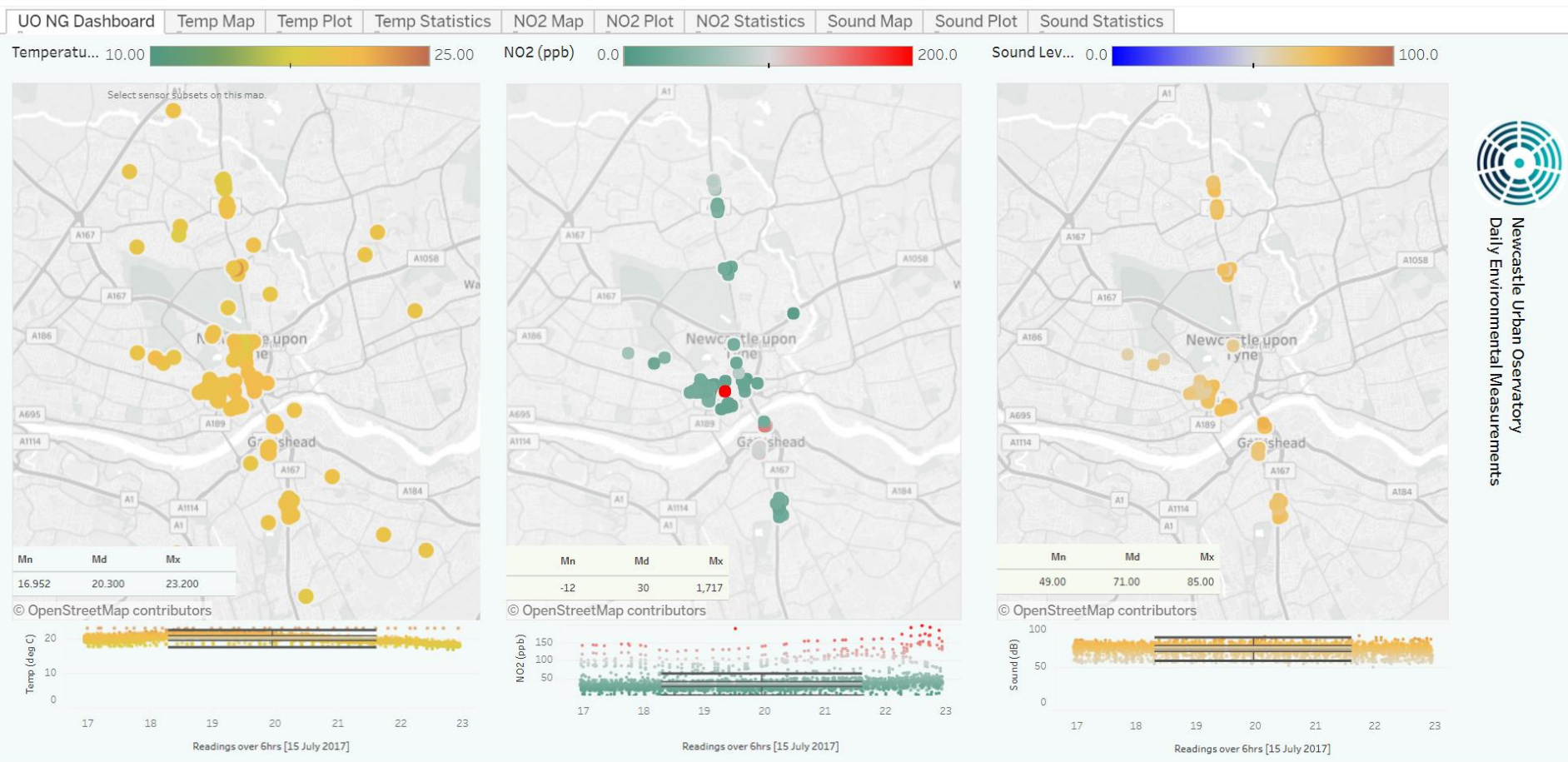
Similar to KNIME a dataflow/pipeline building system.
Commercially licensed software, but free for academic use.

Tableau



Interactive data analytics and visualization tool.

Widely used wysiwig interface, one of the leading toolsets.

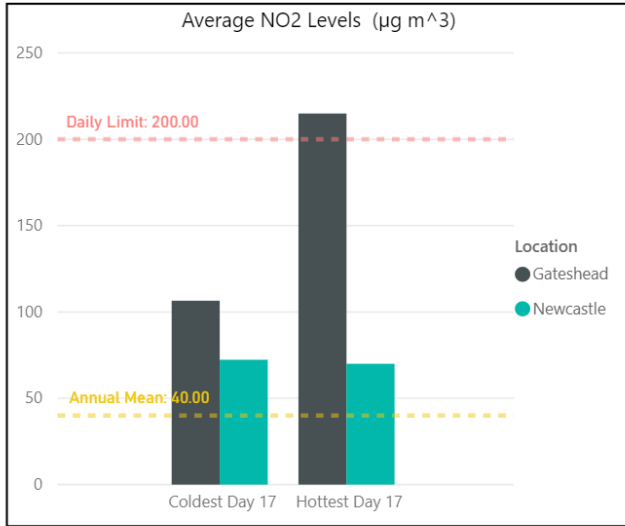
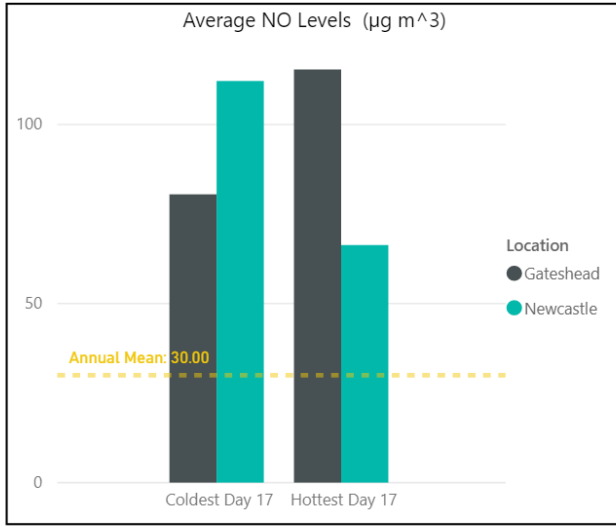
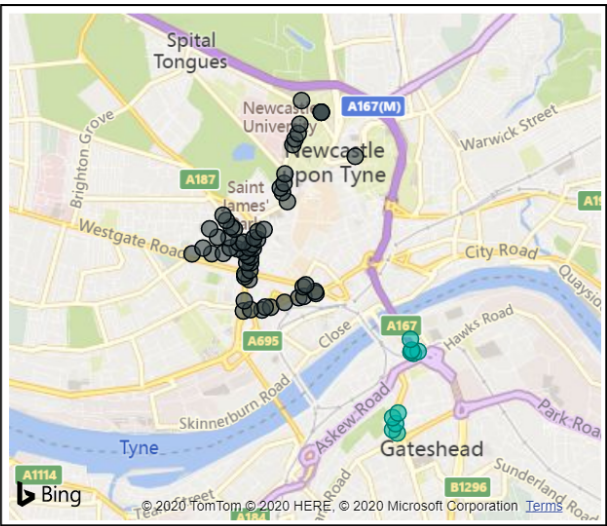


Newcastle Urban Observatory
Daily Environmental Measurements

Power BI

Microsoft's Interactive Data Analytics Tool.
Possibly the leading interactive visualization toolset.

Hottest / Coldest Day of the Year
Newcastle vs. Gateshead



Event	Year	Month	Day
Coldest Day 17	2017	February	17
Hottest Day 17	2017	August	2

Hottest and Coldest Day, on average, according to:
<https://weatherspark.com/y/42189/Average-Weather-in-Newcastle-upon-Tyne-United-Kingdom-Year-Round>

Hourly Limit (not to be exceeded more the 18 times per year)
and Target Annual Mean as specified in DEFRA UK Air Quality targets.

