The British Museum

Transforming a museum to be data-driven using R

Alice Daish
Data Scientist
@alice_data

- Ladies London
August 2016

The British Museum

Set up

Opened 1759 to all 'studious and curious persons'
1st National Public Museum in the World

Today
2nd most visited museum in the world
8 million objects
2 million years of human history

The British Museum

Starting point

Didn't have …

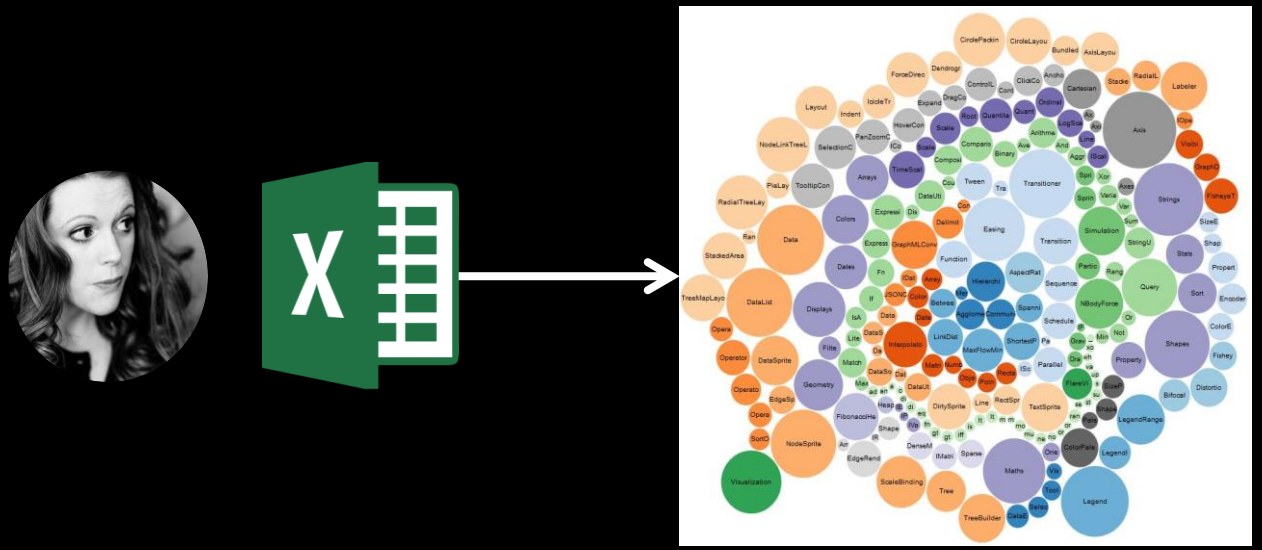
- No list of data sources
- No data access
- No databases
- No data warehouse

Did have…

- R
- Data Scientist
- Big Data: Senior Product manager
- What does "big data" mean to the museum?

The British Museum

Bubbles Envy



D3 Bubble Chart: https://bl.ocks.org/mbostock/4063269

Joe Cheng to the rescue

d3.js bubble chart htmlwidget for R

This R package provides a bubble chart as seen in this Mike Bostock example. It is based on htmlwidgets so it can be used from the R console, RStudio, R Markdown documents, and Shiny applications.

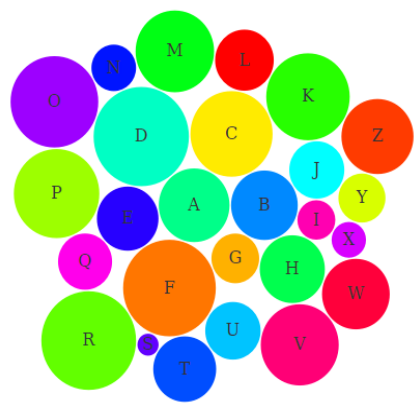
Installation

Use the **devtools** package (`install.packages("devtools")`) to install this package directly from GitHub:

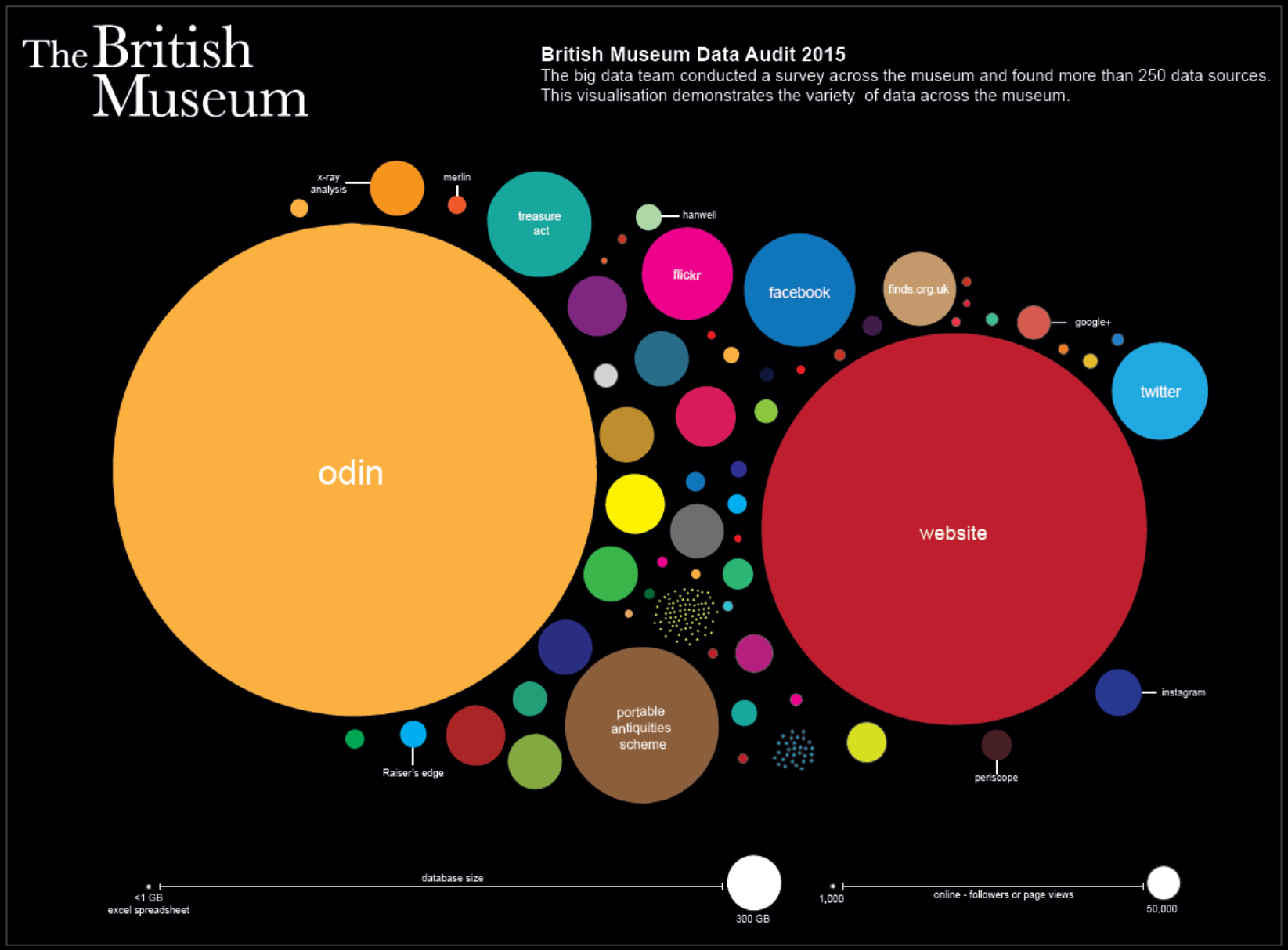
```
devtools::install_github("jcheng5/bubbles")
```

Usage

```
library(bubbles)

bubbles(value = runif(26), label = LETTERS,
  color = rainbow(26, alpha=NULL)[sample(26)]
)
```



https://github.com/jcheng5/bubbles



@alice_data

Package: bubbles

The British Museum

Business Problems = Data Opportunities

We don’t know who our visitors are?
Online = > 9 million
Offline = 6.8 million

We don’t what they do in the museum?

And we don’t know the opportunities to generate revenue?

The British Museum

Business Problems = Data Opportunities

We don't know who our visitors are?
Online = > 9 million
Offline = 6.8 million

We don't what they do in the museum?

And we don't know the opportunities to generate revenue?

"silos" and "wrangling"

data viz

visitor movement

predictive modelling

The British Museum

“silos” and “wrangling”

Multiple visitor data platforms

CSV exports from external platforms

No SQL

Packages: data.table, dplyr

The British Museum

“silos” and “wrangling”

Multiple visitor data platforms

CSV exports from external platforms

No SQL

100’s of columns

print format exports nested by timeslots

Split first and second name

Packages: data.table, dplyr

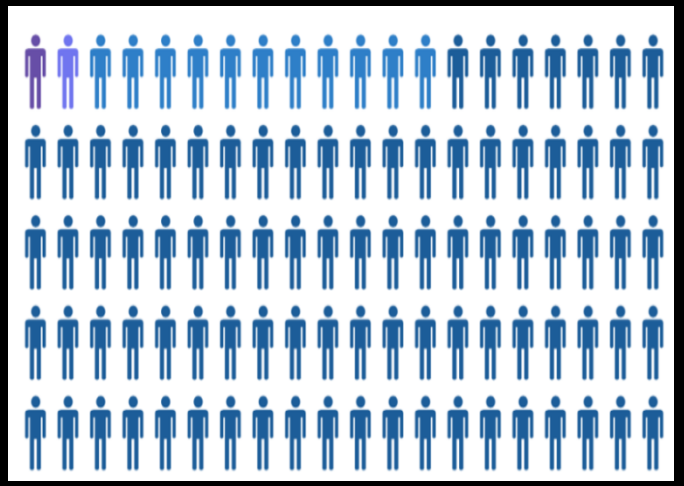
The British Museum

No SQL = data.table

How many visitors are on multiple platforms?

Assess the visitor data siloing

Why?
To improve engagement and access of the museum we need to examine our visitor data.

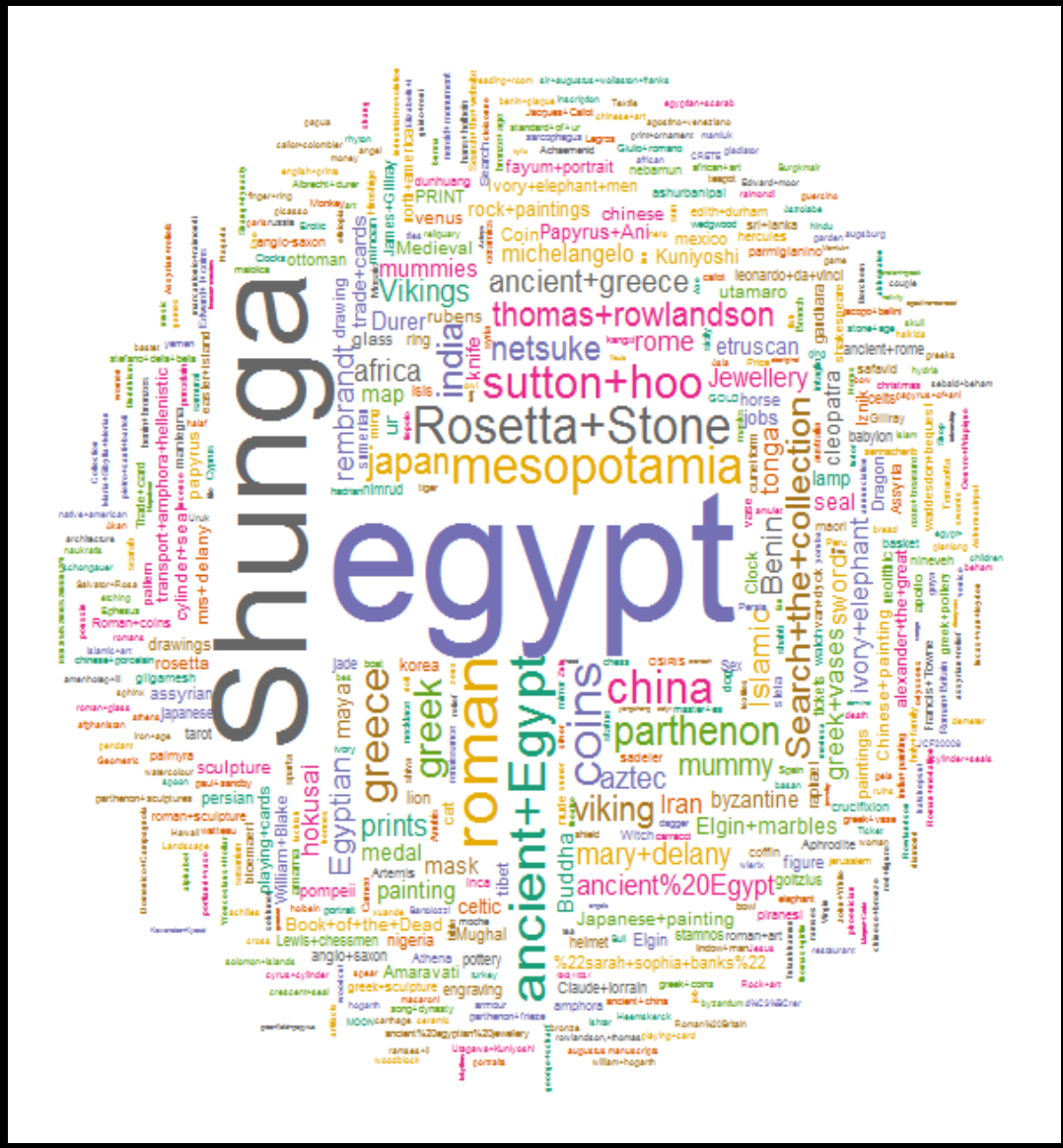


Packages: data.table, dplyr

The British Museum

Top 500 website search

June 2015-16



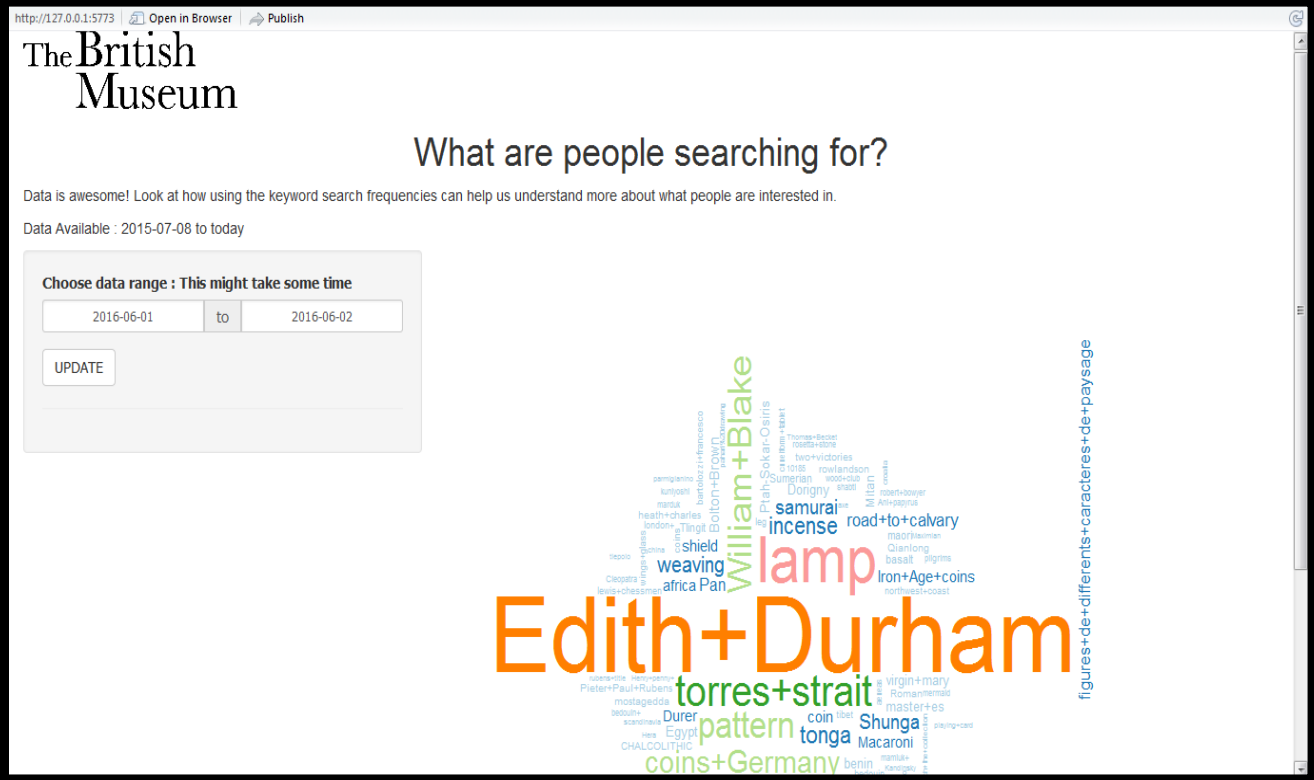
Packages : RSiteCatalyst (Adobe Analytics), WordCloud

@alice_data

Alice Daish. (2016). R KeyWord Search WordCloud: First Release. Zenodo. 10.5281/zenodo.56283

The British Museum

shiny potential

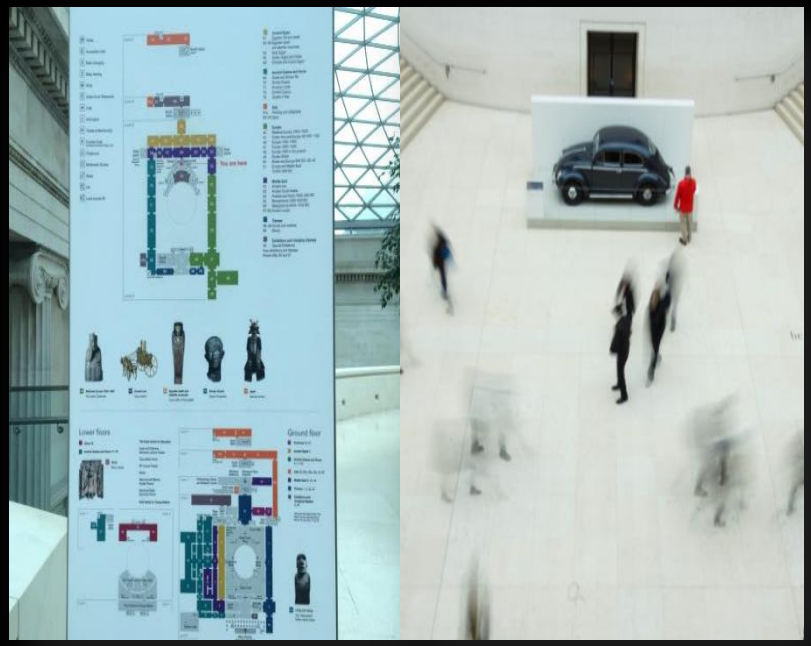


Packages : Shiny, RSiteCatalyst (Adobe Analytics), WordCloud

The British Museum

visitor movement

62 galleries, 3 floors, largest covered public square in Europe with 6.8 million visitors per year.

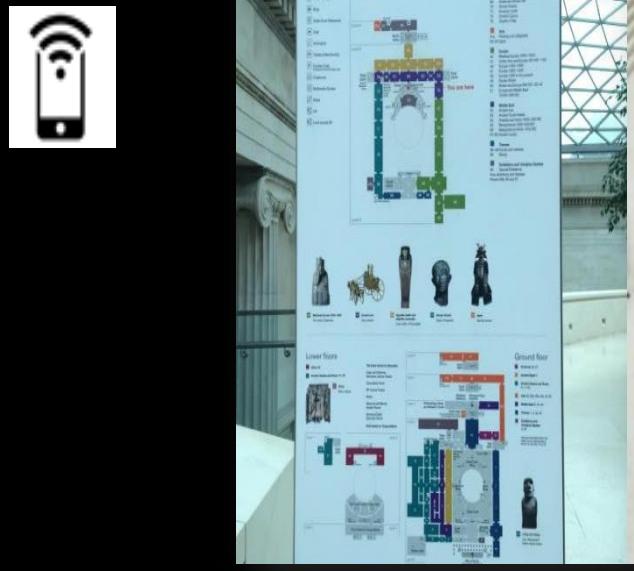


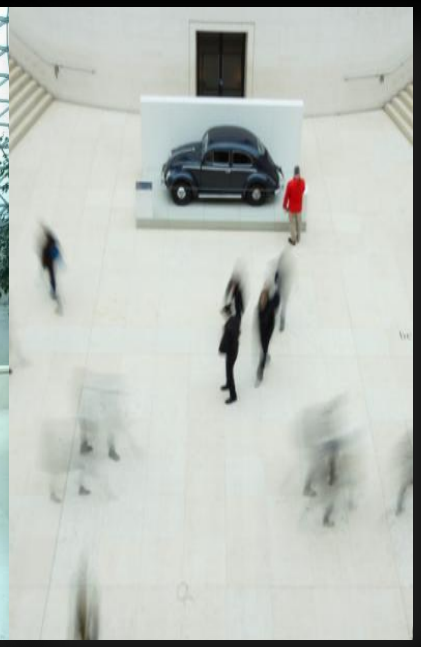
Packages : httr, jsonlite

The British Museum

Visitor Movement

Wi-Fi presence used to sample visitor numbers





1st to use R to connect to CISCO Presence API

Packages : httr, jsonlite

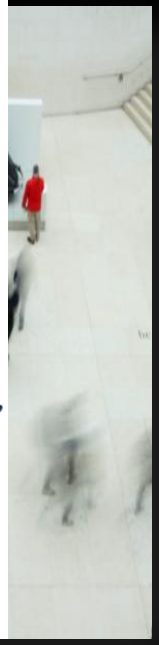
The British Museum

Visitor M

Wi-Fi presence
used to sampl
visitor numbe

1st to use R to connect to CIS Presence API

```r
#-------------CISCO API Connection to collect CISCO data--------------
#DATE : 03/06/2016
#AUTHOR: Alice Daish adaish@britishmusuem.org

#-----Install and Load packages
library(httr)
library(jsonlite)

#Load password and username code file
source("logincisco.R")

#----FIND THE LIST OF SITES-------
sites<-GET("https://cmxcisco.com/api/config/v1/sites",authenticate(user, password))
# gets the URL api content including authroziation

#testing different export formats
str(content(sites)) #see content
sitelist<-content(sites, "text") #collects content as text string
sitelist<-fromJSON(sitelist) #convert to table format from string
head(sitelist) #see the top of the table

#List of site name and siteId
siteId<-cbind(sitelist$aesUidString,sitelist$name)

#EXAMPLE COLLECT HOURLY DATA OF ALL SITE FOR ONE DAY (14/05/2016)
hourdata<-matrix(NA,nrow = 97*1, ncol = 27) #blank matrix
colnames(hourdata)<-c("SiteID","SiteName","Date","0","1","2","3","4","5",
                      "6","7","8","9","10","11","12","13","14","15","16",
                      "17","18","19","20","21","22","23") #Label columns

  hourly<-GET(paste0("https://1nzgy2.cmxcisco.com/api/presence/v1/visitor/hourly?siteId=",
                     siteId[i,1],"&date=2016-05-14"),authenticate(user, password))

  hourdata[i,1]<-siteId[i,1] #ID
  hourdata[i,2]<-siteId[i,2] #Name of site
  hourdata[i,3]<-"2016-05-14"
  hourdata[i,4]<-content(hourly)$`0`
  hourdata[i,5]<-content(hourly)$`1`
  hourdata[i,6]<-content(hourly)$`2`
  hourdata[i,7]<-content(hourly)$`3`
```



@alice_data

Packages : httr, jsonlite

The British Museum

Visitor M

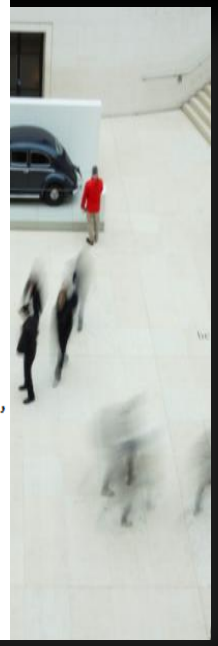
Wi-Fi presence used to sampl visitor numbe

1st to use R to connect to CIS Presence API

```
#-------------CISCO API Connection to collect CISCO data--------------
#DATE : 03/06/2016
#AUTHOR: Alice Daish adaish@britishmusuem.org
```

9-10am	10-11am	11am-12p	12-13pm	13-14pm	14-15pm	15pm-16p	16-17pm	17-18pm	Total
23	155	212	254	174	165	155	141	120	1399
38	76	94	153	120	115	101	158	92	947
26	52	82	92	91	84	80	105	54	666
27	50	68	81	74	65	83	87	98	633
7	23	64	93	68	65	69	85	36	510
11	61	78	61	76	48	61	57	37	490
13	56	45	88	47	27	31	78	69	454
26	40	42	65	56	58	48	47	48	430
14	29	53	111	48	31	35	72	35	428
1	26	69	62	51	64	66	56	27	422
1	39	78	57	52	45	53	52	19	396
0	30	56	57	53	59	59	53	23	390
3	24	56	56	47	55	41	43	15	340
15	18	29	30	34	41	51	36	86	340
14	43	30	34	44	43	36	31	50	325
5	20	48	73	38	36	30	44	27	321
7	39	53	43	38	30	42	39	27	318
3	22	44	54	41	36	39	50	26	315
4	16	39	50	50	38	38	42	36	313
0	30	31	28	43	44	37	38	16	267
6	17	30	53	24	24	31	40	28	253
19	27	20	30	18	22	23	31	38	228
2	14	21	39	37	35	25	20	22	215
8	27	32	23	27	29	25	18	22	211
12	12	15	22	28	27	29	25	37	207
22	38	33	24	22	19	15	16	15	204
11	12	19	33	32	21	25	32	17	202
2	17	26	26	31	26	25	30	12	195
10	33	22	25	10	17	20	29	4	170
2	19	17	29	31	19	20	19	12	168
0	9	23	28	19	17	25	35	11	167
0	15	24	31	13	16	20	35	12	166
11	14	21	33	23	20	18	13	8	161
11	9	27	28	23	23	19	12	6	158
3	19	19	25	15	15	24	18	18	156

```
hourdata[i,7]<-content(hourly)$`3`
```



@alice_data

Packages : httr, jsonlite

The British Museum

Predictive modelling

Can we predict ticket sales for exhibitions?

mixed effect modelling

- data wrangling
- modelling
- prediction

Packages : lme4, lubridate, dplyr, data.table, ggplot2, ggthemes

The British Museum

Predictive modelling

First initial model created – lmer()
Predicted first exhibition sales –predict()
Development continues …

Packages : lme4, lubridate, dplyr, data.table, ggplot2, ggthemes

The British Museum

Future

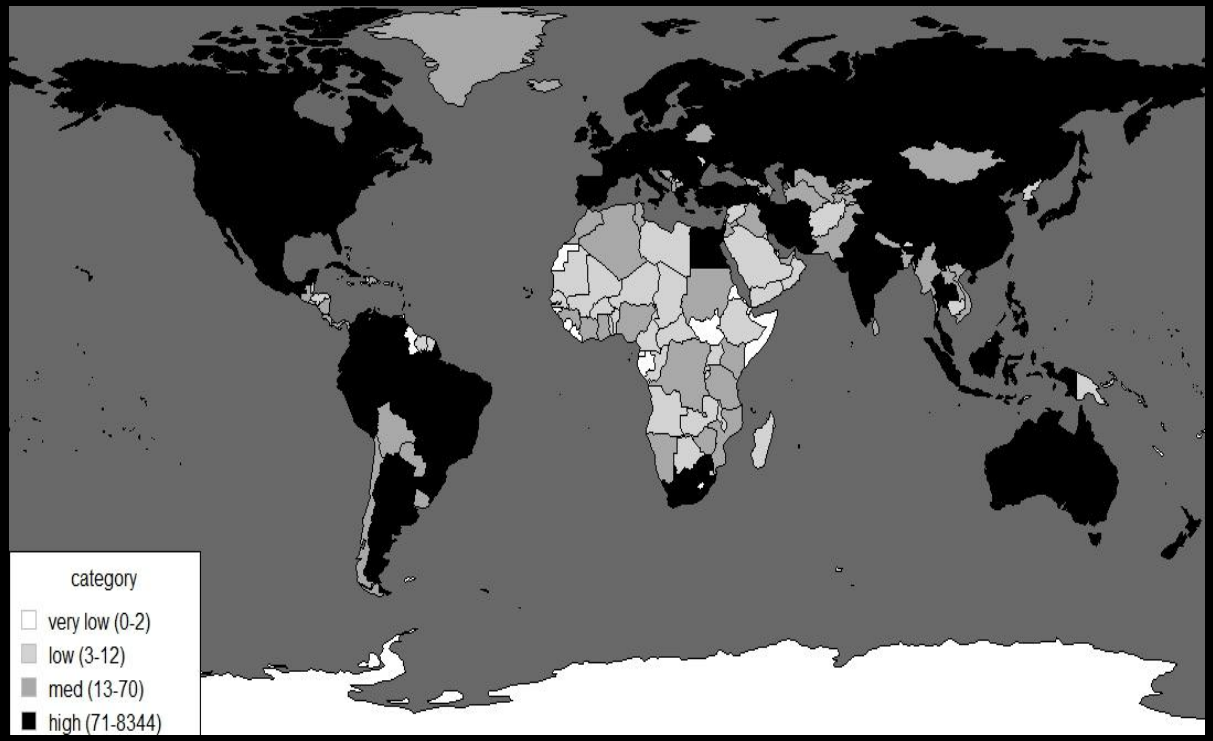
Building a data pipeline including R.
Continued data wrangling of the museums data sources to find insights and value.

Who knows?
Internet of things
e.g. Toilet doors locks, Boilers, Visitor Flow Signs

Machine Learning
e.g. Predicting Visitor Numbers, Optimization, Cognitive Services.

museums
potential

55,000 museums 180 countries



Packages : rworldmap

Data Sources: Museums of the World (2015) Berlin, Boston: De Gruyter Saur. Retrieved 5, Jun 2016, from http://ww.degruyter.com/view/product/448539

The British Museum

Thank you & Questions

adaish@britishmuseum.org
@alice_data

Many thanks to museum departments support and data access, Siorna Ashby, and the R community for their continued support

Symbol - Noun Project: Dara Ullrich, DE

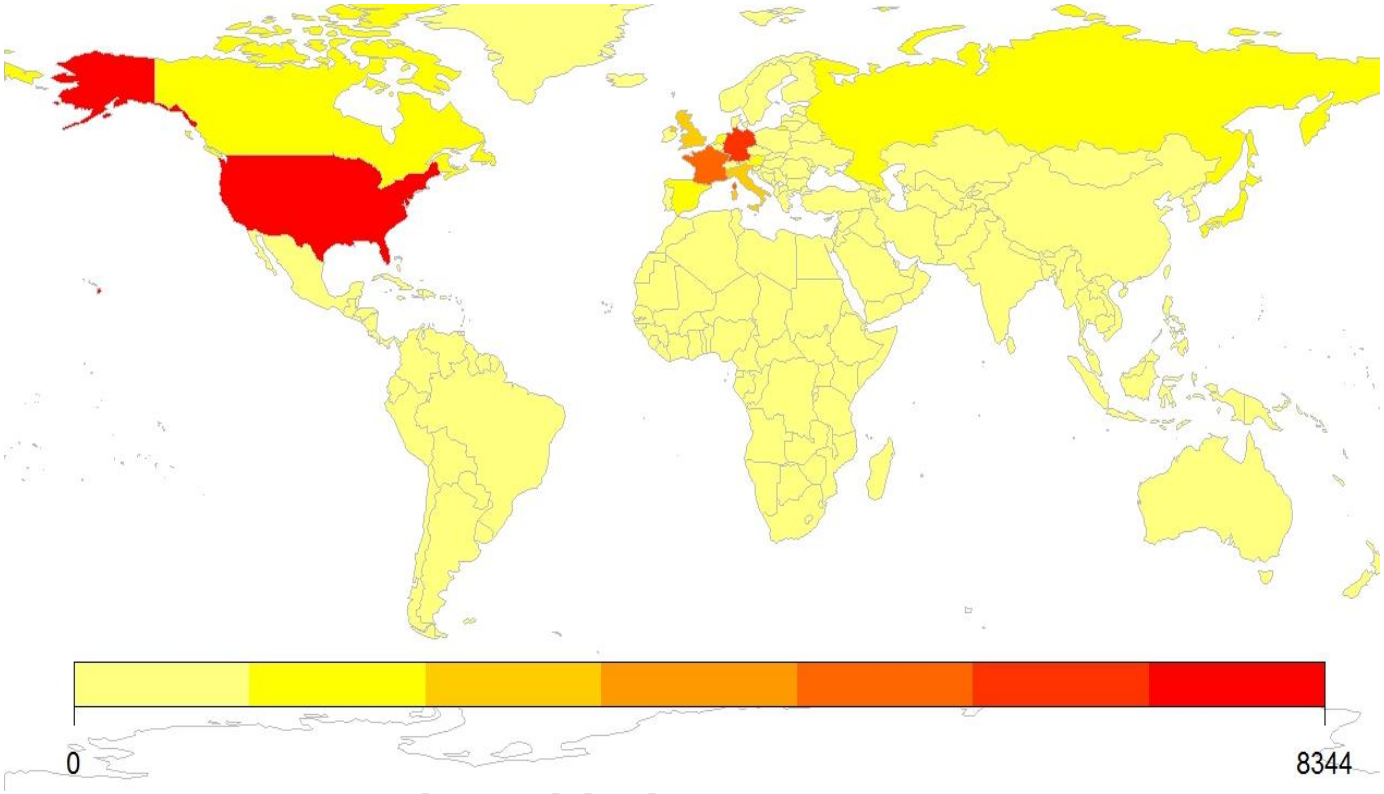
8344
USA

6464
Germany

4842
France

3200
Italy

2943
UK



Countries with the most museums

Data Sources: Museums of the World (2015) Berlin, Boston: De Gruyter Saur. Retrieved 5, Jun 2016, from http://ww.degruyter.com/view/product/448539