

Food Borne Pathogens

Anthony Girard, Timothy Hedspeth, Yutong Li

2022-10-05

Introduction

Motivating question

Literature review

Data Overview

What data are we using

In this toy example I have downloaded data from a random strain of salmonella as to to explore what we can do with this data.

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

##   X.Organism.group      Strain Isolate.identifiers      Serovar
##   "character"          "character"      "character"      "character"
##   Isolate              Create.date      Location      Isolation.source
##   "character"          "character"      "character"      "character"
##   Isolation.type        Min.same        Min.diff        BioSample
##   "character"          "character"      "character"      "character"
##   Assembly              Computed.types
##   "character"          "character"

##   X.Organism.group      Strain
## 1 Salmonella enterica swgs1023
## 2 Salmonella enterica Durban
## 3 Salmonella enterica FAR0085
## 4 Salmonella enterica
## 5 Salmonella enterica
## 6 Salmonella enterica

##   Isolate.identifiers      Serovar
## 1 "SRS569085","swgs1023" Enteritidis
## 2 "Durban","Salmonella enterica from Graham's blood" Enteritidis
## 3 "FAR0085","SRS594015" Enteritidis
## 4 "ASERH121498-sc-2013-02-28T11:34:57Z-1570930","ERS218089" Enteritidis
## 5 "ASERH120613-sc-2013-02-28T11:35:11Z-1570942","ERS218101" Enteritidis
## 6 "ASERH122384-sc-2013-02-28T11:35:18Z-1570948","ERS218107" Enteritidis
##   Isolate      Create.date      Location
## 1 PDT000004258.3 2014-03-15T21:03:16Z      USA:NY
```

```

## 2 PDT000025285.1 2014-04-04T16:50:24Z South Africa: Durban
## 3 PDT000026271.3 2015-02-11T00:08:14Z Pakistan
## 4 PDT000030855.2 2014-05-28T14:43:20Z
## 5 PDT000030857.2 2014-05-28T14:43:20Z
## 6 PDT000030861.3 2014-05-28T14:43:21Z
##      Isolation.source      Isolation.type Min.same Min.diff
## 1      stool      clinical      8      14
## 2      blood      clinical     10     16
## 3 spice mixture for lahori fish environmental/other      8      6
## 4      NA      NA
## 5      NA      NA
## 6      NA      NA
##      BioSample      Assembly      Computed.types
## 1 SAMN02676287 GCA_008015975.1 antigen_formula=9:g,m:-,serotype=Enteritidis
## 2 SAMN02712059 GCA_000612325.1 antigen_formula=9:g,m:-,serotype=Enteritidis
## 3 SAMN02345283 GCA_011084405.1 antigen_formula=9:g,m:-,serotype=Enteritidis
## 4 SAMEA1920593      antigen_formula=9:g,m:-,serotype=Enteritidis
## 5 SAMEA1920604      antigen_formula=9:g,m:-,serotype=Enteritidis
## 6 SAMEA1920611      antigen_formula=9:g,m:-,serotype=Enteritidis

##
##
##      1297
##      Australia
##      13
##      Australia: NSW
##      19
##      Austria
##      14
##      Belgium
##      29
##      Canada
##      92
##      Canada:Prince Edward Island
##      1
##      Chile
##      5
##      Denmark
##      158
##      Finland
##      5
##      Germany
##      201
##      Germany: Bavaria
##      1
##      Greece
##      1
##      Ireland
##      83
##      Italy
##      29
##      Kenya
##      16
##      Kuwait

```

##	1
##	Lithuania
##	1
##	Mauritius
##	20
##	Mexico
##	4
##	Namibia
##	1
##	Netherlands
##	249
##	New Zealand
##	11
##	Northern Ireland
##	5
##	Norway
##	8
##	Pakistan
##	1
##	Poland
##	56
##	Romania
##	1
##	Russia
##	33
##	Senegal
##	1
##	South Africa
##	99
##	South Africa: Durban
##	1
##	South Africa: Eastern cape
##	5
##	South Africa: Eastern Cape
##	233
##	South Africa: Free state
##	4
##	South Africa: Free State
##	131
##	South Africa: Free State, Bloemfontein
##	2
##	South Africa: Gauteng
##	994
##	South Africa: KwaZulu-Natal
##	259
##	South Africa: Limpopo
##	75
##	South Africa: Mpumalanga
##	55
##	South Africa: North West
##	121
##	South Africa: Northern Cape
##	28
##	South Africa: Orange Free State

##		5
##	South Africa: Western cape	
##		1
##	South Africa: Western Cape	
##		566
##	Trinidad and Tobago: Trinidad	
##		11
##	Tunisia	
##		25
##	Uganda	
##		15
##	United Kingdom	
##		115
##	United Kingdom: England	
##		1
##	United Kingdom: France	
##		5
##	United Kingdom: Germany	
##		9
##	United Kingdom: London	
##		92
##	United Kingdom: Midlands and East of England	
##		149
##	United Kingdom: None	
##		120
##	United Kingdom: North of England	
##		182
##	United Kingdom: South of England	
##		200
##	United Kingdom: United Kingdom	
##		5227
##	United Kingdom: United Kingdom, United Kingdom	
##		16
##	United Kingdom: Wales	
##		9
##	USA	
##		5512
##	USA: SC	
##		1
##	USA:AL	
##		11
##	USA:AR	
##		1
##	USA:CA	
##		45
##	USA:CO	
##		5
##	USA:CT	
##		9
##	USA:DE	
##		38
##	USA:FL	
##		11
##	USA:GA	

##	28
##	USA:HI
##	2
##	USA:IA
##	8
##	USA:IL
##	14
##	USA:IN
##	1
##	USA:KS
##	3
##	USA:KY
##	4
##	USA:MA
##	25
##	USA:MD
##	24
##	USA:MI
##	12
##	USA:MN
##	32
##	USA:MO
##	8
##	USA:MS
##	7
##	USA:NC
##	106
##	USA:ND
##	2
##	USA:NE
##	3
##	USA:NH
##	5
##	USA:NJ
##	35
##	USA:NM
##	1
##	USA:NY
##	31
##	USA:OH
##	58
##	USA:OK
##	2
##	USA:OR
##	3
##	USA:PA
##	20
##	USA:PR
##	2
##	USA:RI
##	6
##	USA:SC
##	41
##	USA:TN

```
## 17
## USA:TX
## 13
## USA:VA
## 14
## USA:WA
## 21
## USA:WI
## 28
## USA:WV
## 8
```

```
## [1] 17328
```

All the data reads in as character. There are 17,328 in this data set alone hailing from many countries around the world

Missing Data

Exploratory Analysis

Conclusion