```
library(dplyr)

rladies_global %>%
  filter(city == ‘London')
```

R-LADIES LONDON LIGHTENING TALKS

My R Journey

Caroline Kovacs

UNIVERSITY OF PORTSMOUTH

My R Journey…

Caroline Kovacs, Centre for Healthcare Modelling & Informatics

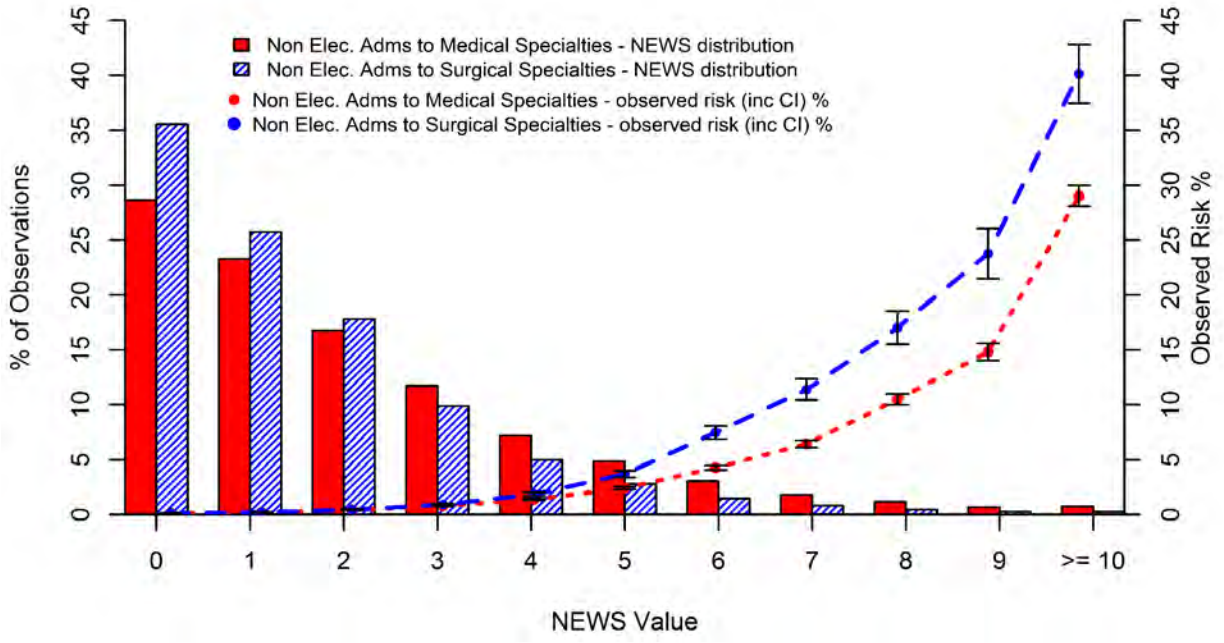
R Ladies Meet-Up November 2017

My R Journey…

- Clinical outcome data research…
  - Vital Signs Observations
  - National Early Warning Score (NEWS) & surgical patients
  - Operating Theatre data
- Useful functions:
  - str()
  - hist()
  - summary()
  - plot()
- Take-away..
  - RStudio
  - R Notebooks

UNIVERSITY OF PORTSMOUTH

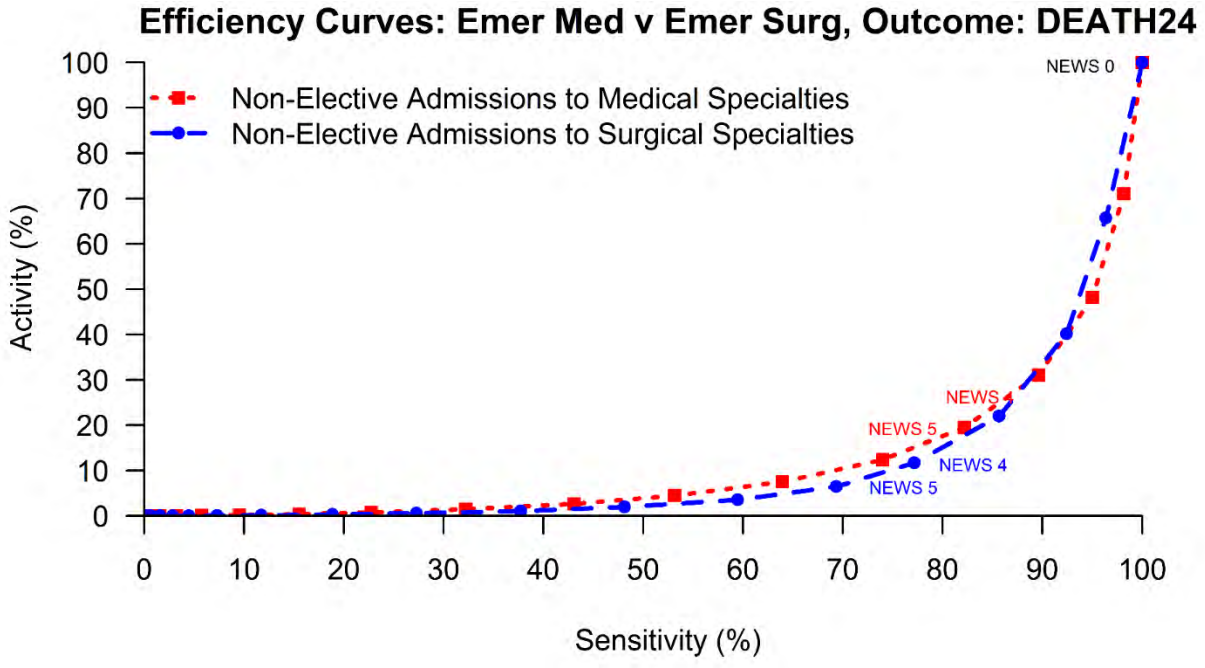
Sample plot…
- NEWS distribution v Risk



Kovacs, C., Jarvis, S. W., Prytherch, D. R., Meredith, P., Schmidt, P. E., Briggs, J. S., & Smith, G. B. (2016). Comparison of the National Early Warning Score in non-elective medical and surgical patients, 1385–1393. http://doi.org/10.1002/bjs.10267

UNIVERSITY OF PORTSMOUTH

Sample plot…



Kovacs, C., Jarvis, S. W., Prytherch, D. R., Meredith, P., Schmidt, P. E., Briggs, J. S., & Smith, G. B. (2016). Comparison of the National Early Warning Score in non-elective medical and surgical patients, 1385–1393. http://doi.org/10.1002/bjs.10267

UNIVERSITY OF PORTSMOUTH

UNIVERSITY OF PORTSMOUTH

Thank-you

caroline.kovacs@port.ac.uk

Saving Excel Sheets from R

Erle Holgersen

Saving Excel Sheets from R

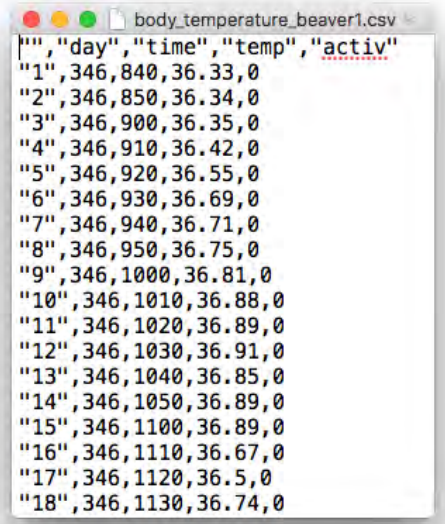
Erle Holgersen

November 28th, 2017

Basic Approach

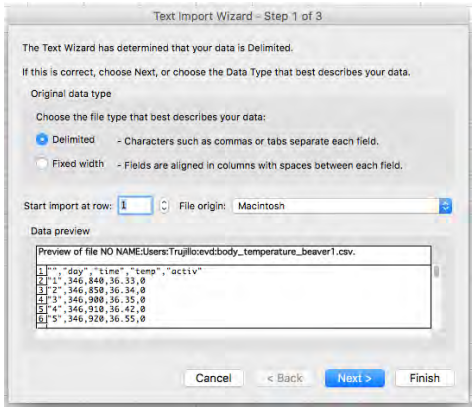
- Write CSV file from R, import into Excel

```
write.csv(
    beaver1,
    'body_temperature_beaver1.csv'
    );
```



Why Try Something Else?

- Data typically requires formatting after importing it
- If you know the end user will be using Excel, you can automate the process



	A	B	C	D	E
1		day	time	temp	activ
2	1	346	840	36.33	0
3	2	346	850	36.34	0
4	3	346	900	36.35	0
5	4	346	910	36.42	0
6	5	346	920	36.55	0
7	6	346	930	36.69	0
8	7	346	940	36.71	0
9	8	346	950	36.75	0
10	9	346	1000	36.81	0
11	10	346	1010	36.88	0
12	11	346	1020	36.89	0
13	12	346	1030	36.91	0
14	13	346	1040	36.85	0

Saving Excel Files

- Can use openxlsx package (or alternatives)
- Allows for saving directly to formatted xlsx sheets

```
library(openxlsx);

write.xlsx(
    beaver1,
    file = 'body_temperature_beaver1.xlsx'
    );
```

Making a Fancier Workbook

```
wb <- createWorkbook();
addWorksheet(wb, sheetName = 'Beaver 1');

writeData(
    wb, sheet = 'Beaver 1',
    x = t(c('Day', 'Time', 'Temperature (°C)', 'Active')),
    colNames = FALSE
    );

writeData(
    wb, sheet = 'Beaver 1',
    x = beaver1, colNames = FALSE,
    startRow = 2);

# bold headers
addStyle(wb, sheet = 'Beaver 1', rows = 1, cols = 1:4,
    style = createStyle(textDecoration = 'bold')
    );

saveWorkbook(wb, 'beavers.xlsx');
```

	A	B	C	D
1	Day	Time	Temperature (°C)	Active
2	346	840	36,33	0
3	346	850	36,34	0
4	346	900	36,35	0
5	346	910	36,42	0
6	346	920	36,55	0
7	346	930	36,69	0
8	346	940	36,71	0
9	346	950	36,75	0
10	346	1000	36,81	0
11	346	1010	36,88	0
12	346	1020	36,89	0
13	346	1030	36,91	0
14	346	1040	36,85	0
15	346	1050	36,89	0
16	346	1100	36,89	0
17	346	1110	36,67	0
18	346	1120	36,5	0
19	346	1130	36,74	0

Conclusion

- Non-programmers love Excel sheets
- Lots of packages available for saving directly to Excel sheets
- Depending on your use-case, it could be worth investing time in formatting output

Summer Storm Loss Models

Jessica Turner

Building a Summer Storm Loss Model in R

Jessica Turner
@j_k_turner

Catastrophe Loss Modelling 101

Property Locations

HAZARD

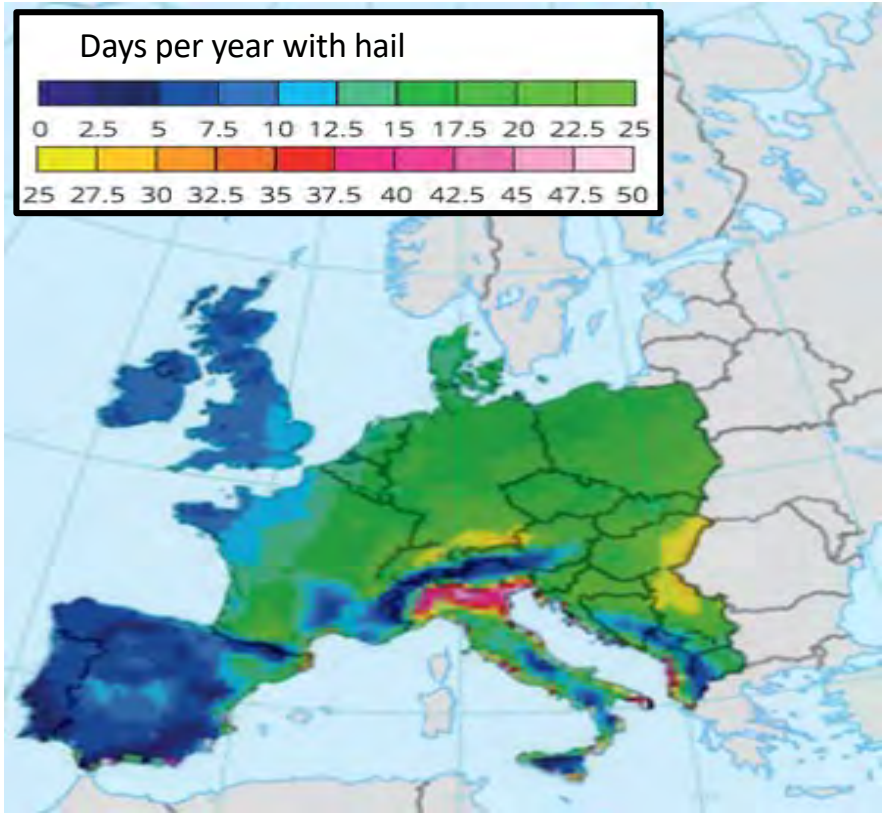
DAMAGE

POLICY TERMS

Loss

Summer Storms

- Convective-driven
  - Causes heavy rain, lightening, hail, wind and tornadoes
  - Strength measured by CAPE (kJ)
- Convection likes
  - Hot land surface
  - Moist/dry air boundaries
  - Mountain foothills
- Relatively weak in the UK




Thank you!
NATIONAL GEOGRAPHIC
Photograph by Maximilian Conrad | National Geographic Traveler Photo Contest 2014
© Copyright Maximilian Conrad. All rights reserved.

Global Health Strategy & Shiny

Ainize Cidoncha

Xtrategy

Enhancing Global Health strategy planning using an
R Shiny iterative tool

Ainize Cidoncha

Challenge yourself

Innovate

Impact

CHALLENGE YOURSELF

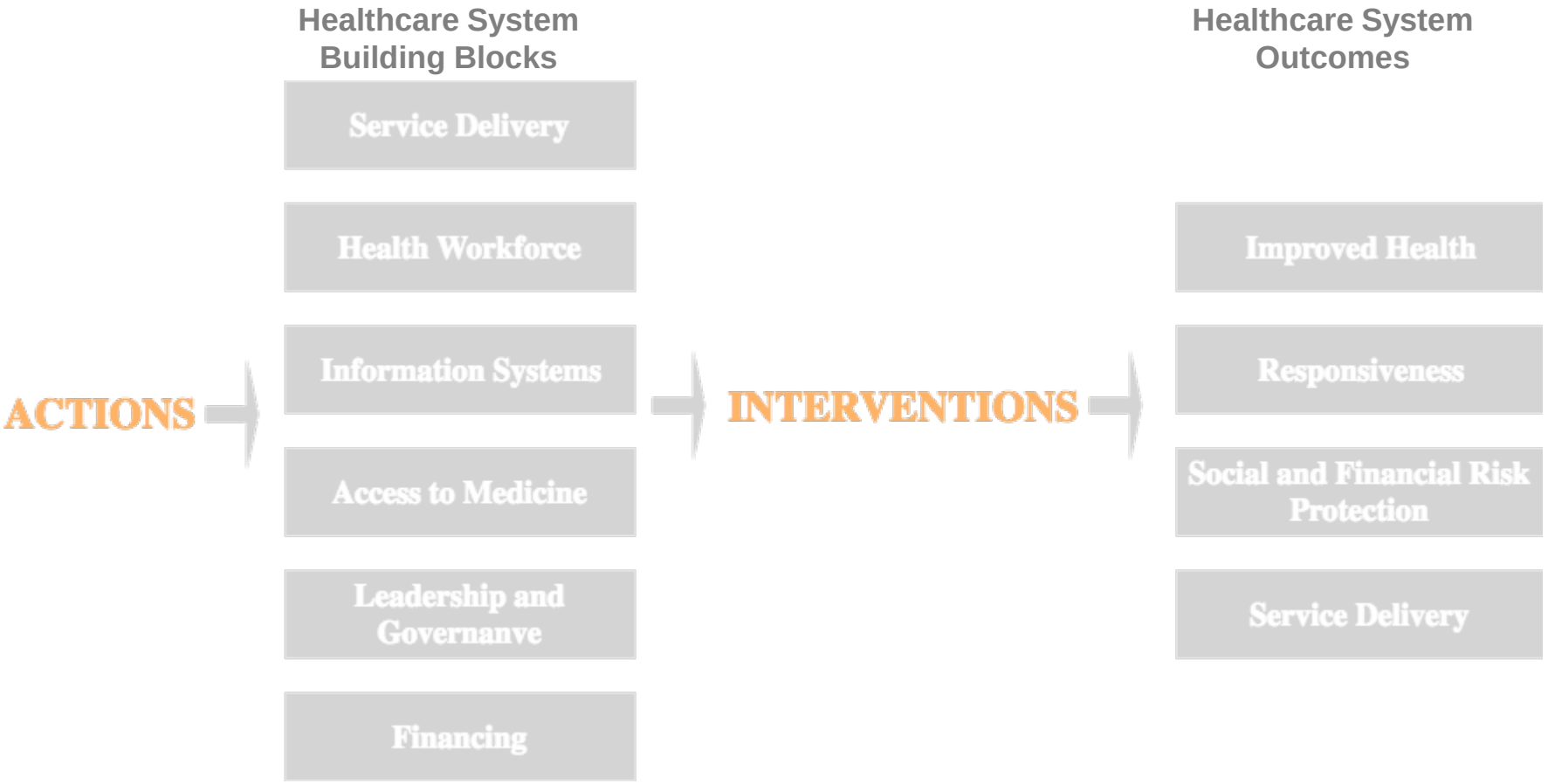
Which are

the most suitable **strategies**

for **scaling up** the care

of **Chagas** disease

in **Bolivia**?


Healthcare System Building Blocks
Service Delivery
Health Workforce
Information Systems
Access to Medicine
Leadership and Governanve
Financing
ACTIONS
INTERVENTIONS
Healthcare System Outcomes
Improved Health
Responsiveness
Social and Financial Risk Protection
Service Delivery

Ranking of Interventions

Intervention	Outcomes	Complexity
Screening people donating blood banks	0.23	0.06
Follow-up blood safety management in Hospitals	0.27	0.18
Briefings at Blood banks about CH/TX	0.29	0.09
Improvement and construction of housing	0.32	0.24
Entimological Surveillance	0.31	0.24
Following up newborns from seropositive mothers	0.41	0.21
Identify children in schools by applying a risk survey	0.20	0.11
Fumigation of houses based on infestation results	0.27	0.12
Post-sprayed evaluation	0.27	0.19
Selective spraying	0.20	0.17
E MOCHA: App for detection of infected houses	0.31	0.18
Community Education and material provision	0.40	0.13

Ranking of Actions

Description	# Int.	%
Financing through increment expenditure of the MoH	37	69%
Introduce in the POAs (Annual municipal operational plans)	32	59%
Strength the general knowledge about Chagas disease	31	57%
Management and organization of health centers	30	56%
Lead the community involvement supporting interventions	30	56%
Policy for the chagas management at a national level	30	56%
Develop an intervention plan	29	54%
Monitore and quality control of the data entry	29	54%
Supply of inputs and essential drugs	29	54%
Training and supporting plan to primary health care HW	26	48%
Integrated software	25	46%

INNOVATE

IMPACT

Outcomes and ongoing projects

Bolivia	Completed (May 15-Jan 16)
Catalonia	Completed (Jan 17-Jun 17)
Colombia	In Progress (Jun 17)
USA	Planned (Sep 17)

TO BE CONTINUED…

THANK YOU

Ainize Cidoncha
ainizecm@gmail.com
@ainize_cidoncha

Cross Tabulations in R

Suzanne Fox

Cross-Tabs – Why are they interesting ?

AKA : Cross-Tabulations/Contingency Tables/Pivot Tables

- First job in 80's as a Market Research analyst, cross-tabs were the only analytical tool. That company grew from a £1 million pound business then to a £1 billion business today. Software companies selling cross-tab software were successful.

- Cross-tabs are a super-quick way to look for distributions, patterns, trends, and to compare sub-groups within the data. The confusion matrix is a good example of a simple but very useful cross-tab.

- When I started in Data Science I was amazed that I couldn't find libraries in R or Python that gave me easy-to-create cross-tabs. Creating a cross-tab shouldn't need to have a bespoke script for each dataset, it's a generalisable pattern.

- Ability to construct Pivot tables in Excel is a widely asked for skill, but (In my opinion) Pivot tables can be quite labour intensive to create, and they *don't work with tidyverse data*.

Cross-Tabs in R – An Example

Let's look at a toy example of 30 people asked about shopping habits. We know Age, Gender, Number of Shopping Trips and whether they bought Bread, Milk and/or Eggs.

Query *– I want to know whether Men or Women make more shopping trips*

I can cross-tab TripCount by Gender. BUT to get useful information I need to bin (make summary groups) of the TripCount, and look at the mean.

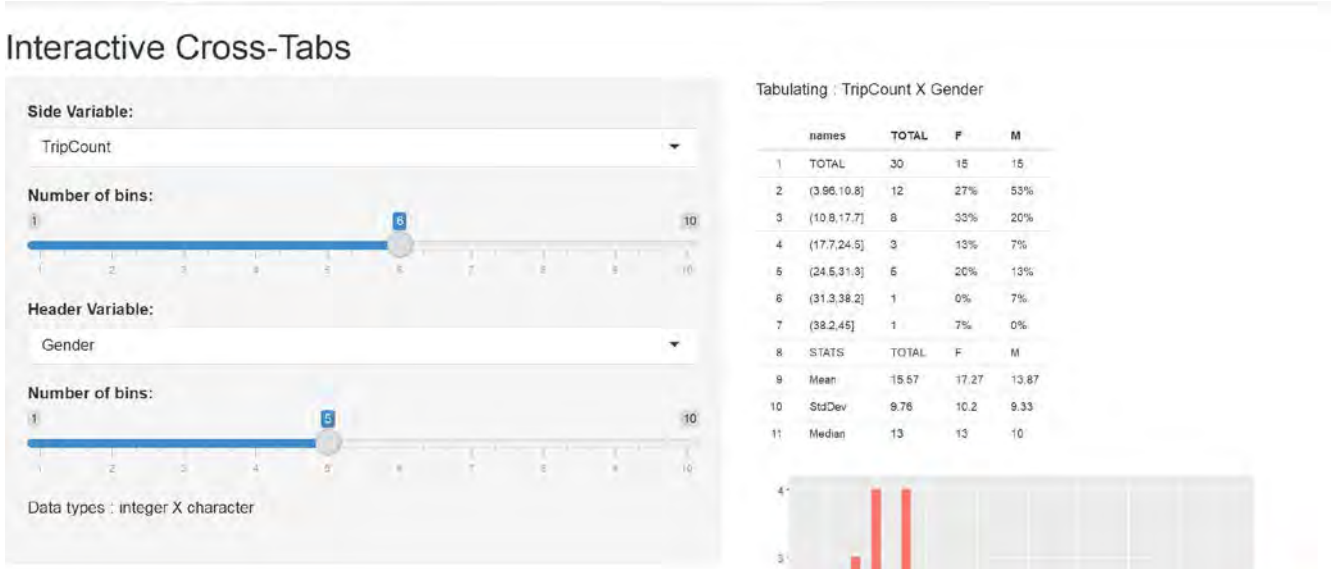
I want –

With the ***minimum amount of code***
Totals as first column not last
Column percents not counts
Stats automatically for numerics
Bins automatically for numerics

ID	Age	Gender	TripCount	Bread	Milk	Eggs
001	25	M	30	0	1	1
002	30	F	28	1	1	1
003	26	M	32	1	1	0
004	32	F	25	1	1	1
005	40	M	25	0	1	1
006	45	F	20	1	1	1
007	18	M	6	0	1	0
008	37	F	26			
009	42	M	22			
010	45	F	18			
011	48	M	16			
012	52	F	12			
013	54	M	8			
014	56	F	13			
015	58	M	10			
016	61	F	13			

```
> crosstab(Data.In, "TripCount", "Gender", 5, 0, 0)
[1] "Data : data.frame"
[1] "Side var : TripCount : integer"
[1] "Head var : Gender : character"
[1] "Time taken : 0.149999999999999 secs"
          names TOTAL     F     M
1         TOTAL    30    15    15
2  (3.96,12.2]    14   40%   53%
3  (12.2,20.4]     8   33%   20%
4  (20.4,28.6]     5   20%   13%
5  (28.6,36.8]     2    0%   13%
6    (36.8,45]     1    7%    0%
7         STATS TOTAL     F     M
8          Mean 15.57 17.27 13.87
9        StdDev  9.76  10.2  9.33
10       Median    13    13    10
```

Cross-Tabs in R – Shiny app for any dataset



Replicates a Pivot table in Excel but with all the automation that I want built in AND I can have appropriate charts without any extra work.

This app works with any data file, you just need to change the file specification – one line of code.

Cross-Tabs in R – Tidy data is even better

What isn't easy to do in Excel or R is make a cross-tab of many variables at the same time, or deal with questions like Q4 where there might be more than 1 answer, for instance if someone buys Bread, Milk and Eggs. If we move to the tidyverse these problems are much easier to solve.

Tidy Data format

ID	Variable	Data
001	Age	25
001	Gender	M
001	Q4	Eggs
001	Q4	Milk
001	TripCount	30
002	Age	30
002	Gender	F
002	Q4	Bread
002	Q4	Eggs
002	Q4	Milk
002	TripCount	28
003	Age	26
003	Gender	M
003	Q4	Bread
003	Q4	Milk
003	TripCount	32
004	Age	32

If the data is tidy, the value of any cell in the cross-tab can be obtained from counting the intersection of two sets of data –

From the toy example –

Set 1 - Side : Variable == "Q4" & Data == "Eggs"
Set 2 - Header : Variable == "Gender" & Data == "M"

So you can use simple dplyr statements to get the data sets from the tidy data, and then just count the records in the intersection set where the Id field has common values in both sets.

```
SideVars <- c("Gender","Q4")
HeaderVars <- c("Gender")

crosstab(data, SideVars, HeaderVars, 0, 0, 0)
```

Set 2

	Side	TOTAL	myCol	myCol1
1	:		Gender : F	Gender : M
2	: TOTAL	30	15	15
3	Gender : F	15	15	0
4	Gender : M	15	0	15
5	Q4 : Bread	17	11	6
6	Q4 : Eggs	13	7	6
7	Q4 : Milk	24	13	11

Set 1

Cross-Tabs in R – Summary

My motivation for developing this is because it's what I'm used to and I have missed this functionality, which to me seems quick, easy to understand and useful.

Have Cross-tabs just been "forgotten" as an EDA/introductory analysis tool or are there better methods these days ?

I'm planning to work on this and develop the Shiny app. Is it generally useful in which case I'll take longer and make it more robust, or is it niche ?

Webscraping for Deals

Ela Osterberger

Dress for less

Ela Osterberger
Director of Analytics at Deliveroo

I like dresses.
I am cheap.

I mainly shop on ASOS because I hate fitting rooms and it's affordable.

...however not affordable enough for me.

I want sales!

R Action plan

1

I have a list of items I want to buy

That's me!

Scrape information from the website

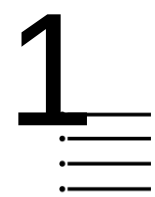
rvest

Find items that are new on sale

If statements

Send myself a shopping list

sendmailR
pander



```
urls <- c(      "http://www.asos.com/blah1”,

          "http://www.asos.com/blah2”,

          "http://www.asos.com/blah3”,
              ...)
```

For loop to
look at all urls

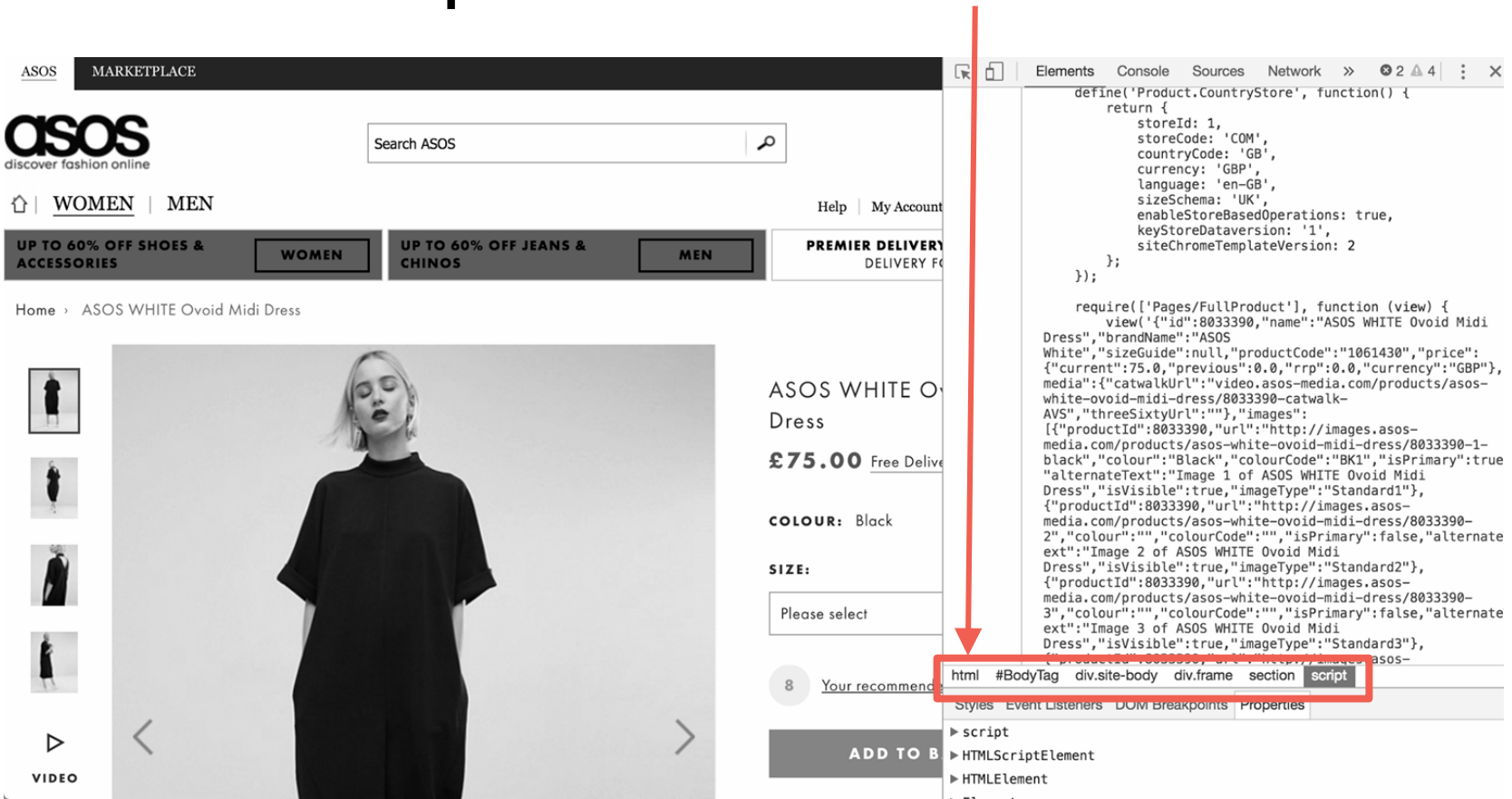
Find the right
part of the html

asos <- **html**(urls[i])

raw <- asos %>% html_nodes("**#BodyTag div.site-body div.frame section script**") %>% html_text()

2

#BodyTag div.site-body div.frame section script




```
asos <- html(urls[i])

raw <- asos %>% html_nodes("#BodyTag div.site-body div.frame
section script") %>% html_text()

string <-
toString(raw[4])
segm <- substr(string,regexpr("current",p),regexpr("current",p)+60)
pictures[i] <-
substr(segm,regexpr("url",segm)+6,regexpr("colour",segm)-4)
```

Some crazy regular expressions

3

```
flag_vector <- rep("NA",length(current_prices))
for (i in seq_along(current_prices)){
      if (current_prices[i]<previous_prices[i])
      flag_vector[i]=1
      else flag_vector[i]=0
  }
table <-
data.frame(c(urls[which(flag_vector>0)]),
      c(previous_prices[which(flag_vector>0)]),
      c(pictures[which(flag_vector>0)]))
colnames(table)= <- c("Url","NEW Price","Pic")
```

Find items whose price has dropped

Build table with info on these items

```
sendmail_options(smtpServer="ASPMX.L.GOOGLE.COM")
from <- "<ela.osterberger@gmail.com>"
to <- "<ela.osterberger@gmail.com>"
subject <- "Let's shop!"
body <- mime_part(paste('<html> <head>
      <html> <p><b> There are new items on sale!
      </b></p></head><body>
      <pre>', paste(pander_return(pander(table[,c(1,3)],
style="multiline")), collapse = '\n'), '</pre>
      </body></html>'))
sendmail(from,to,subject,body)
```

This will build a table

Send!

Let’s do it!

Live Demo