R in Power BI

SQL Saturday #759

Taras Kaduk

2018/05/05

Agenda

- What is R
- R inside Power BI
- Why do you even need it?

Slides and stuff

https://github.com/taraskaduk/r-in-power-bi

~~What is love?~~ What is R?

More importantly, what can it do?

- data transform
- stat analysis
- machine learning / predictive modeling
- web apps
- reports, presentations
- websites & books

What can R do?

Data transformation

```
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

What can R do?

Data transformation

```
iris_gather <- iris %>%
    gather("key", "size", -Species)

head(iris_gather)
```

```
##   Species          key size
## 1  setosa Sepal.Length  5.1
## 2  setosa Sepal.Length  4.9
## 3  setosa Sepal.Length  4.7
## 4  setosa Sepal.Length  4.6
## 5  setosa Sepal.Length  5.0
## 6  setosa Sepal.Length  5.4
```

What can R do?

Data transformation

```
iris_gather2 <- iris_gather %>%
    separate(col = key, into = c('part', 'dimension'))

head(iris_gather2)
```

```
##   Species  part dimension size
## 1  setosa Sepal    Length  5.1
## 2  setosa Sepal    Length  4.9
## 3  setosa Sepal    Length  4.7
## 4  setosa Sepal    Length  4.6
## 5  setosa Sepal    Length  5.0
## 6  setosa Sepal    Length  5.4
```

What can R do?

Data transformation

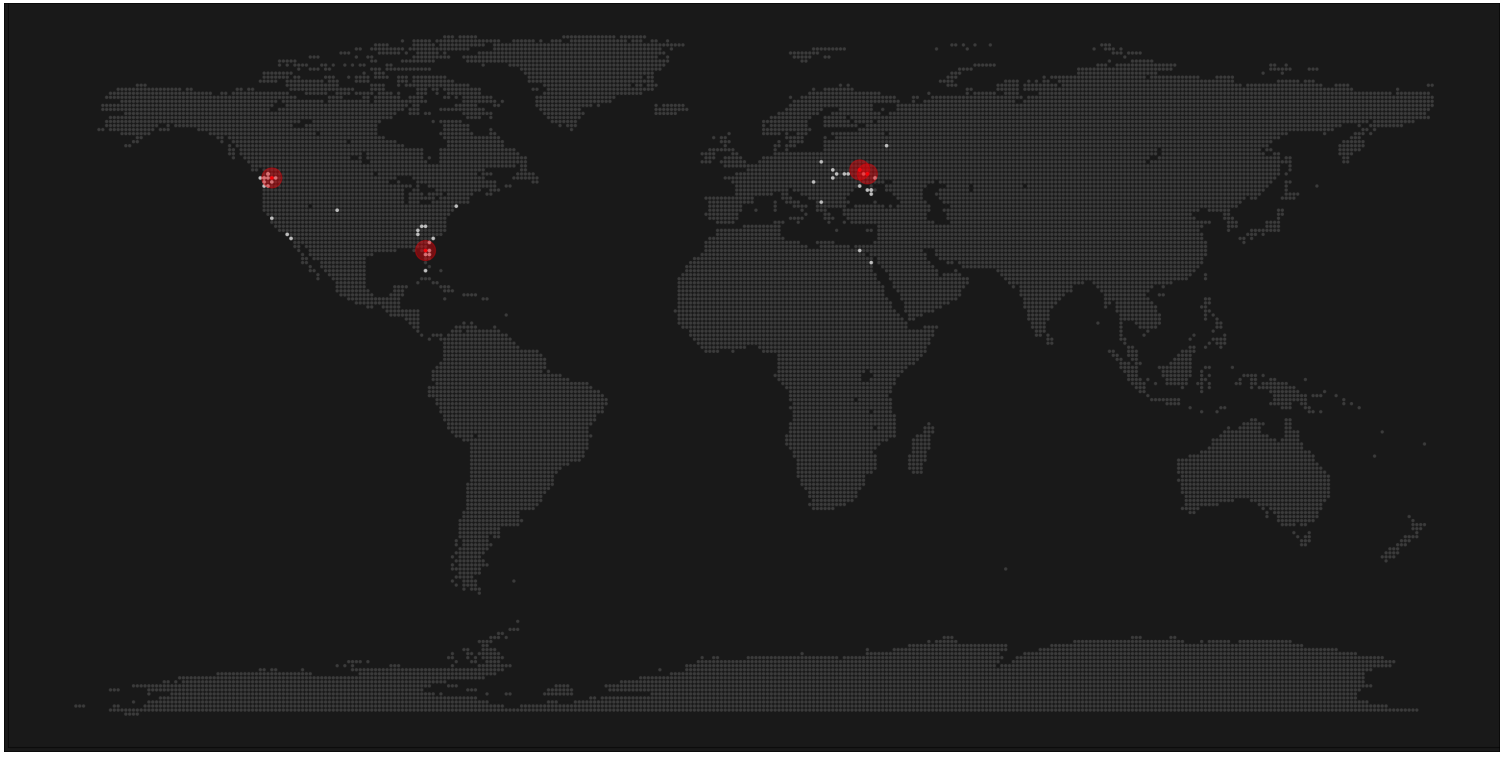
```
iris_summary <- iris_gather2 %>%
    group_by(Species) %>%
    summarise(size_avg = mean(size),
              size_sd = sd(size),
              size_med = median(size))

head(iris_summary)
```

```
## # A tibble: 3 x 4
##   Species    size_avg size_sd size_med
##   <fct>         <dbl>   <dbl>    <dbl>
## 1 setosa         2.54    1.85     2.10
## 2 versicolor     3.57    1.76     3.30
## 3 virginica      4.28    1.92     4.15
```

What can R do? Data visualization

What can R do?

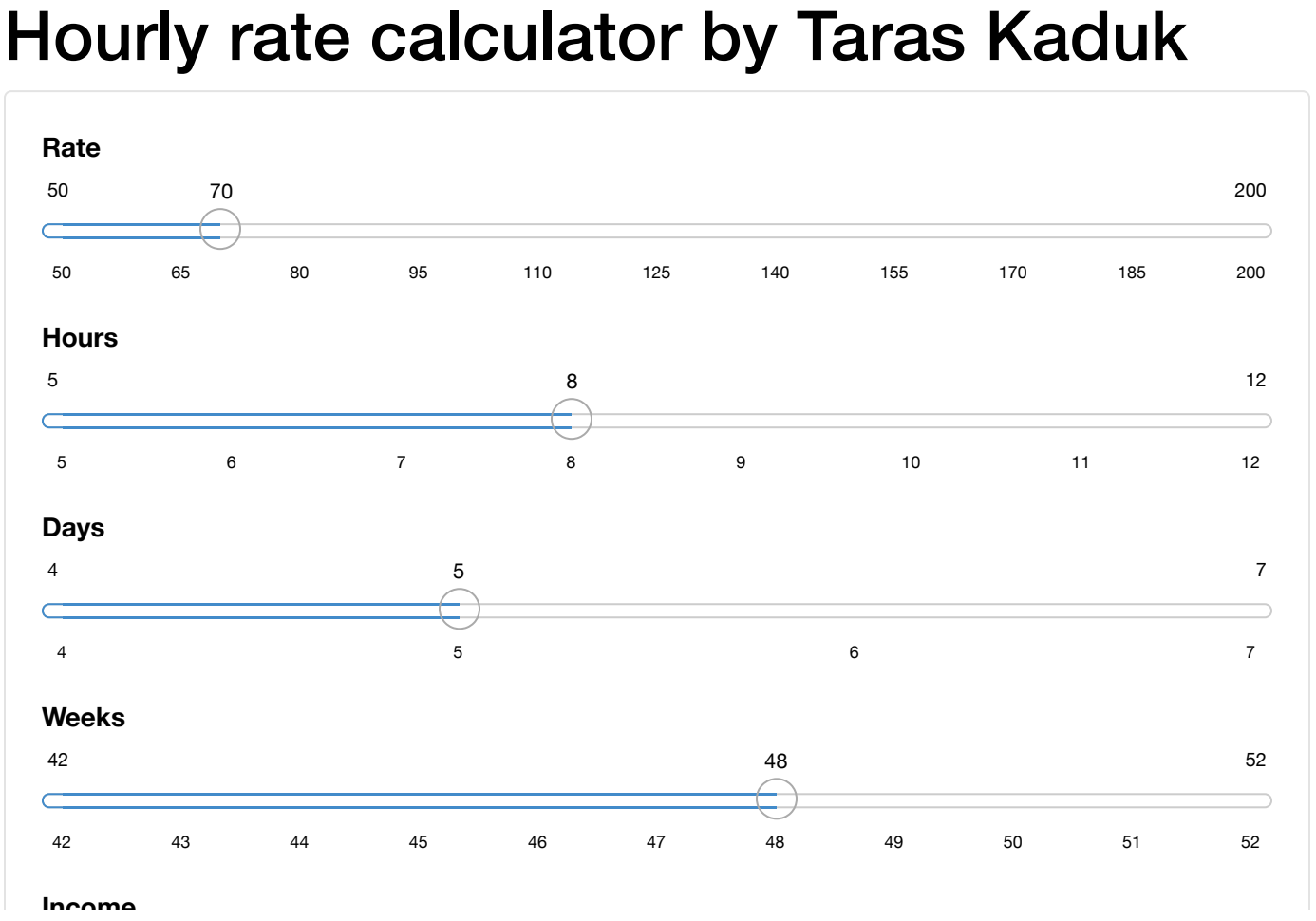
Data visualization



https://taraskaduk.com/2017/11/26/pixel-maps/

What can R do? Web apps (A.k.a Shiny)

https://taraskaduk.shinyapps.io/rate/



What can R do?

So much more!

- reports
- presentations (including this one!)
- websites
- books

What can R do? Bread and butter

> *"The best thing about R is that it was written by statisticians. The worst thing about R is that it was written by statisticians."* ~ **Bob Cowgill** (probably) [1]

1 - Why has R, despite quirks, been so successful? - Revolution Analytics

What can R do?

Statistical analysis and Machine Learning

- perform statistical analysis
- create, train and test models on your laptop (`caret`, `modelr` packages, etc)
- reach to Keras and TensorFlow from within R
- et cetera et cetera

R in Power BI

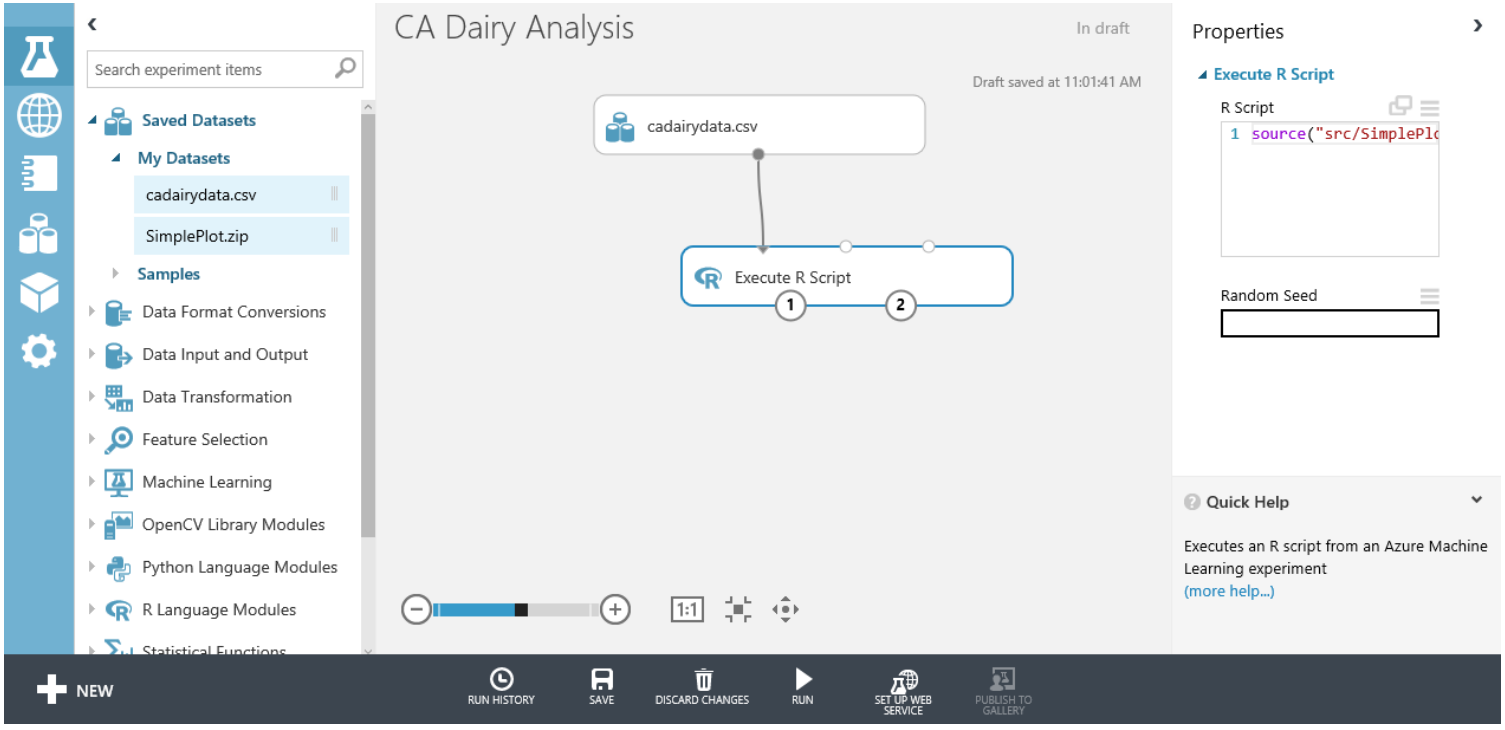
R in Microsoft

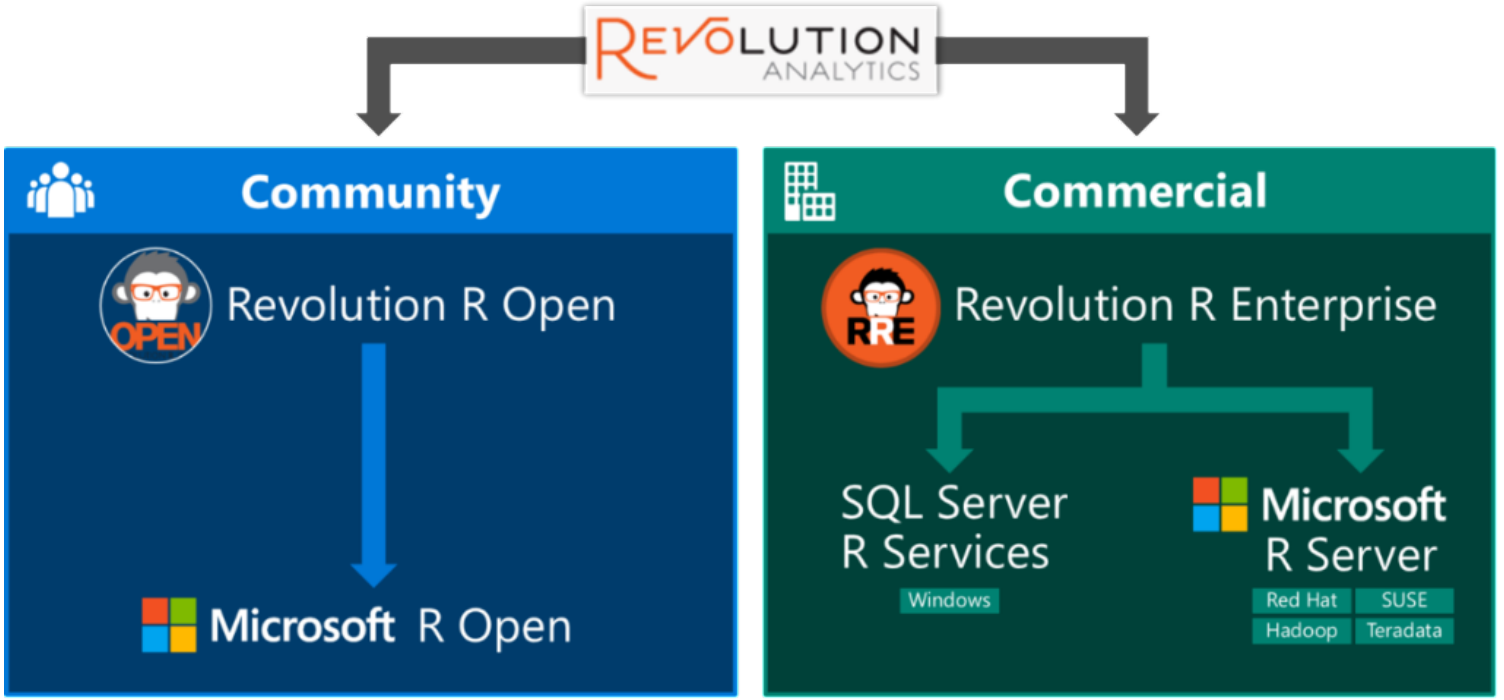
What can you do with R inside of MS products?

- Run R on SQL Server.
- Run R code in Azure ML
- Use Microsoft R Server for paralel processing

https://microsoft.github.io/sql-ml-tutorials/R/rentalprediction/step/3.html

```
-- Stored procedure that trains and generates an R model using the re
DROP PROCEDURE IF EXISTS generate_rental_rx_model;
go
CREATE PROCEDURE generate_rental_rx_model (@trained_model varbinary(m
AS
BEGIN
    EXECUTE sp_execute_external_script
      @language = N'R'
    , @script = N'
        require("RevoScaleR");


            rental_train_data$Holiday = factor(rental_train_data$Holi
            rental_train_data$Snow = factor(rental_train_data$Snow);
            rental_train_data$WeekDay = factor(rental_train_data$Week

        #Create a dtree model and train it using the training data se
        model_dtree <- rxDTree(RentalCount ~ Month + Day + WeekDay +
        #Before saving the model to the DB table, we need to serializ
        trained_model <- as.raw(serialize(model_dtree, connection=NUL

    , @input_data_1 = N'select "RentalCount", "Year", "Month", "Day",
    , @input_data_1_name = N'rental_train_data'
    , @params = N'@trained_model varbinary(max) OUTPUT'
    , @trained_model = @trained_model OUTPUT;
END;
```


CA Dairy Analysis
In draft
Draft saved at 11:01:41 AM
Search experiment items
Saved Datasets
My Datasets
cadairydata.csv
SimplePlot.zip
Samples
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
cadairydata.csv
Execute R Script
1
2
Properties
Execute R Script
R Script
1 source("src/SimplePlo
Random Seed
Quick Help
Executes an R script from an Azure Machine Learning experiment
(more help...)
1:1
NEW
RUN HISTORY
SAVE
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


REVOLUTION ANALYTICS
Community
Revolution R Open
OPEN
Microsoft R Open
Commercial
Revolution R Enterprise
RRE
SQL Server
R Services
Windows
Microsoft
R Server
Red Hat
SUSE
Hadoop
Teradata

Introducing Microsoft R Server

No In-Memory Limitations
Efficient Data Storage Formats

Distributed computation using parallelized algorithms
In-Hadoop & In-Teradata Analysis

Platform longevity
Commercial support and services that Open Source R can't offer

Powerful IDE & Strong Integration
Multi-Platform Scoring

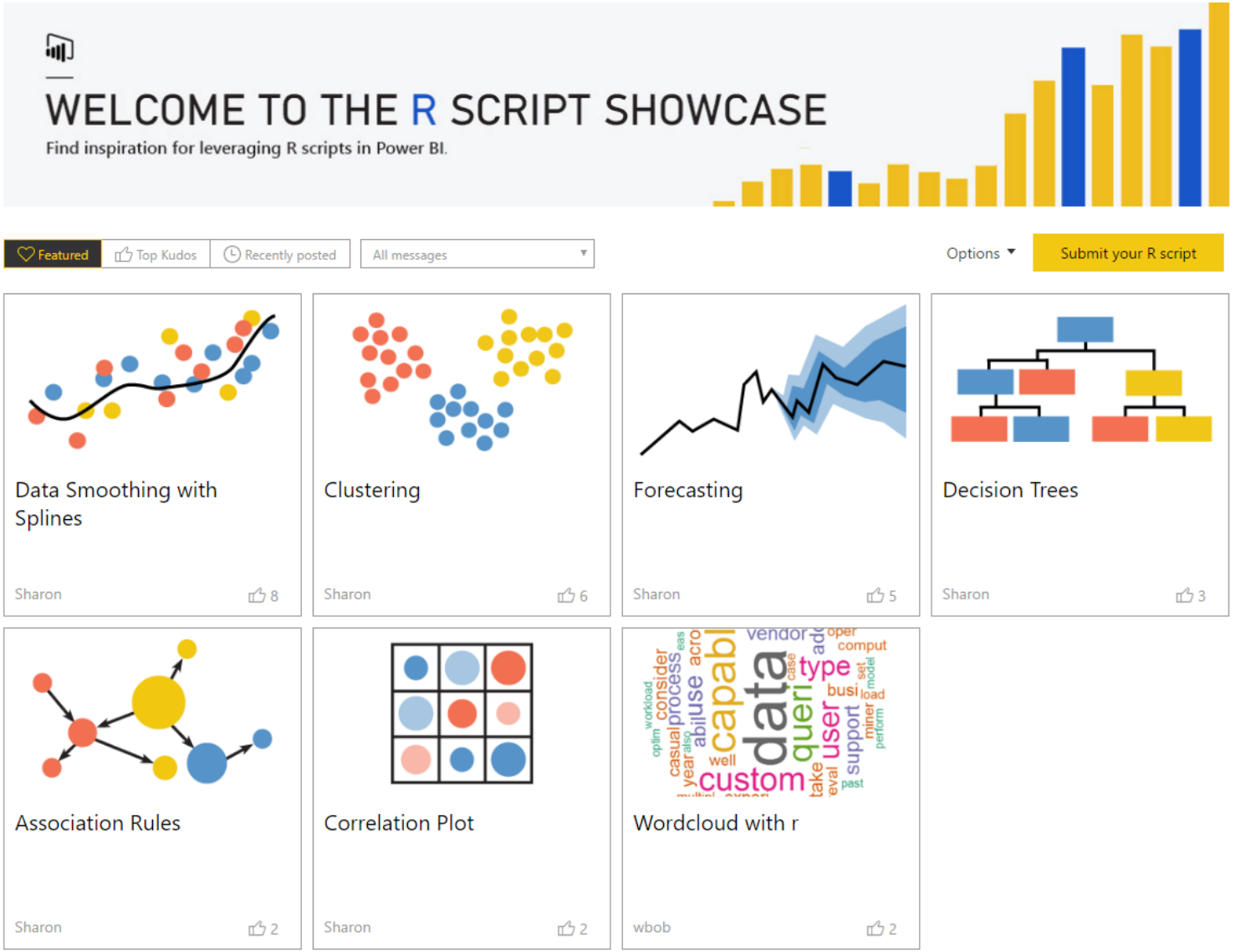
Web services-based integration platform
Application interoperability

R in Power BI

That's cool, but what can I do with R in Power BI?

- custom R visuals
- scripts
- visuals

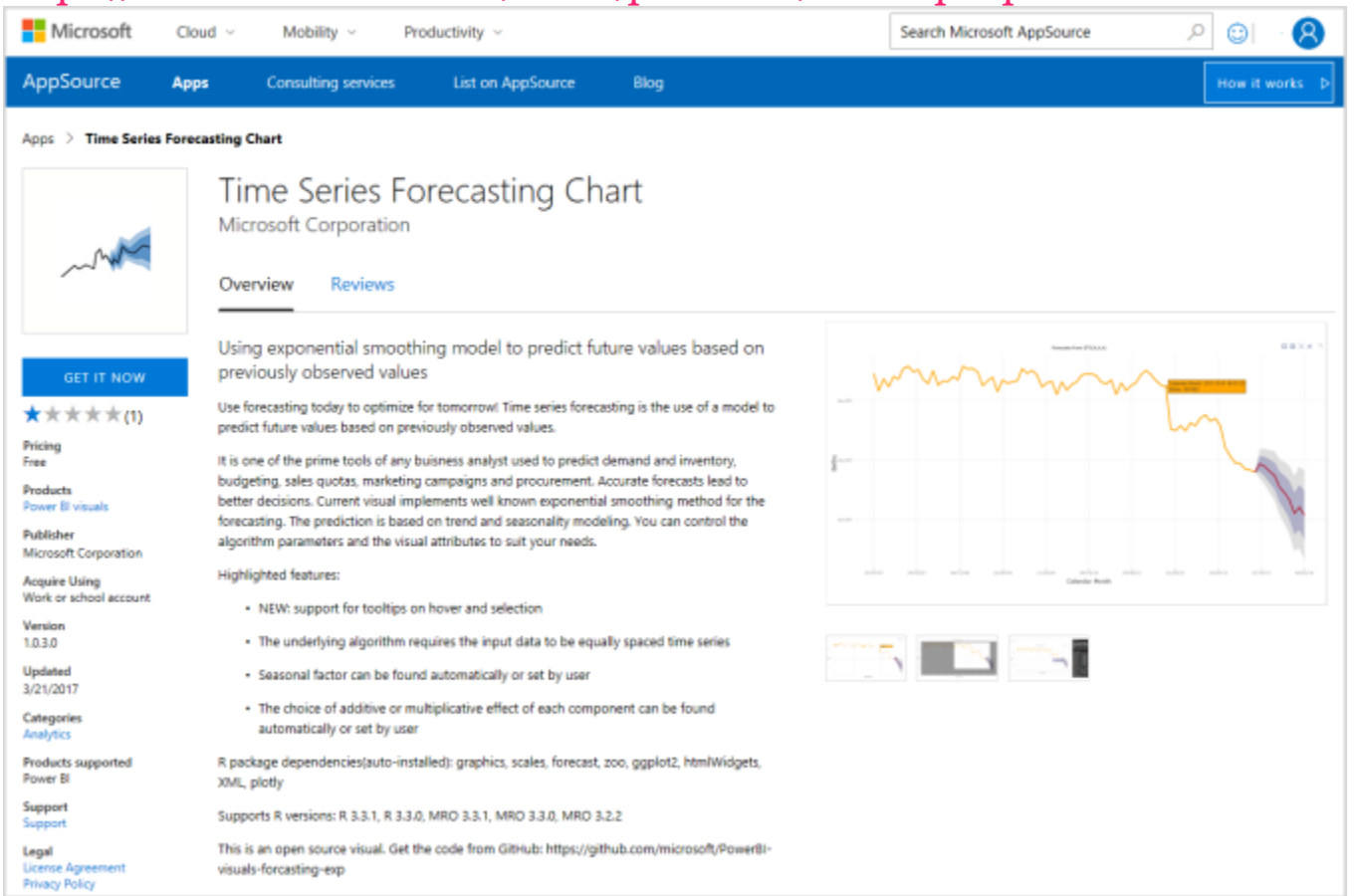
Custom R visuals



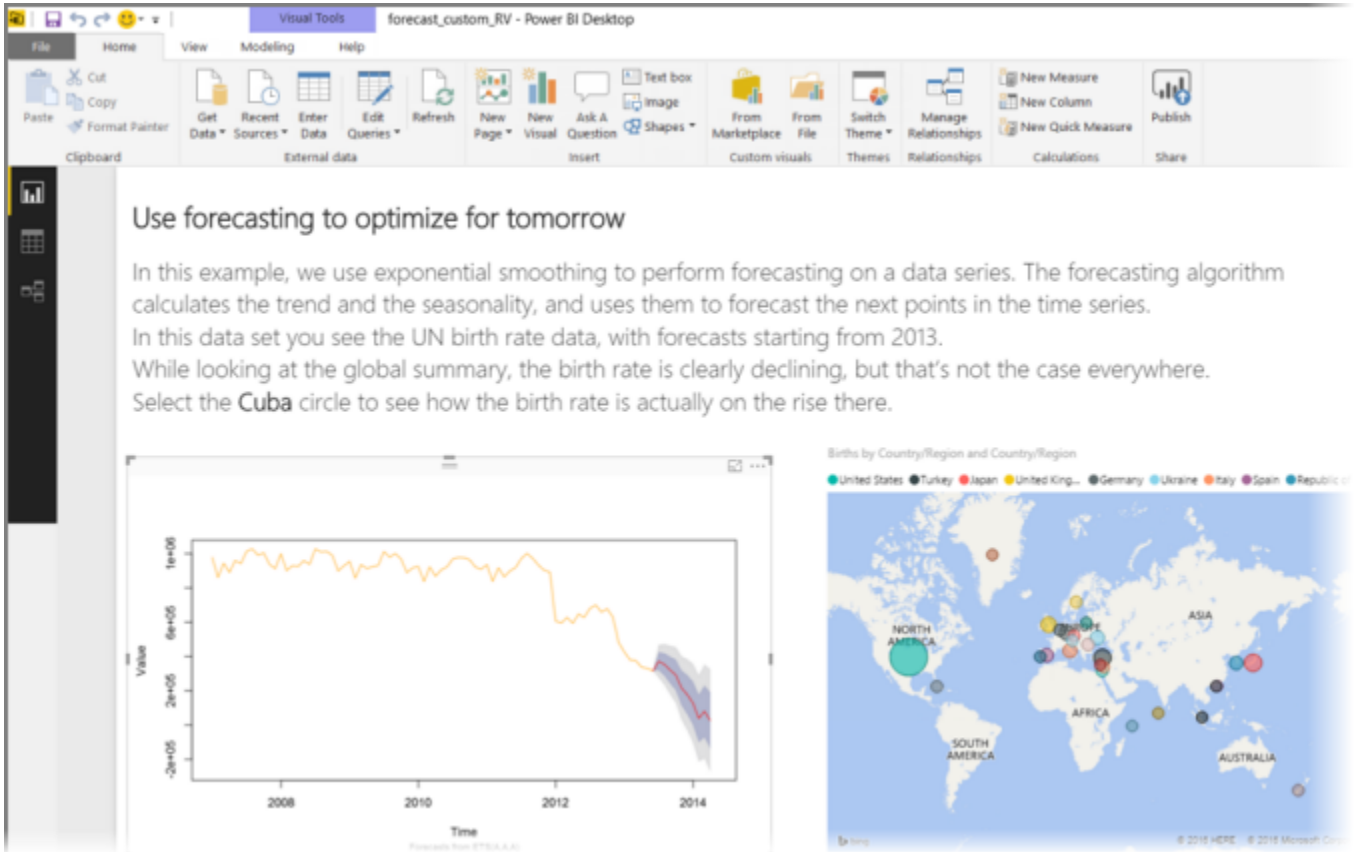
http://community.powerbi.com/t5/R-Script-Showcase/bd-p/RVisuals

Custom R visuals

https://docs.microsoft.com/en-us/power-bi/desktop-r-powered-custom-visuals



Custom R visuals



Custom R Visuals

https://github.com/Microsoft/PowerBI-visuals-forcasting-exp

```
ets_params = list(Automatic="Z",Multiplicative="M",Additive="A",Nor
if(frequency(timeSeries) == 1)
  seasonType = "None"
deModel = paste(ets_params[[errorType]],ets_params[[trendType]],ets

if(sum(deModel==c("ANM","ZMA","MMA","AZM","AMZ","AMM","AMA","AMN","
  deModel = "ZZZ"

fit = ets(timeSeries, model=deModel,damped=damped)
```

Custom R Visuals

https://github.com/Microsoft/PowerBI-visuals-forcasting-exp

```
    p1a <- p1a + labs (title = pbiInfo, caption = NULL) + theme_bw()
    p1a <- p1a + xlab(labTime) + ylab(labValue)
    p1a <- p1a + scale_x_continuous(breaks = seq(1,length(predictions
    p1a <- p1a +  theme(axis.text.x  = element_text(angle = getAngle)
                                                    hjust=1, size = s
                        axis.text.y  = element_text(vjust = 0.5, size
                        plot.title  = element_text(hjust = 0.5, size
                        axis.title=element_text(size =  sizeLabel),
                        axis.text=element_text(size =  sizeTicks),
                        panel.border = element_blank())
```

Custom R visuals in Power BI are well thought-out and overall legit. Use them

R Scripts

R scripts

- When Power Query can't do the job

- When Power Query is slow and clunky

R scripts

This

```
let
    Source = Web.Page(
                Web.Contents(
                "https://github.com/jennybc/gapminder/blob/master/inst/extdat
                )),
    Data = Source{0}[Data],
    col_types = Table.TransformColumnTypes(Data,{
                                    {"", type text},
                                    {"country", type text},
                                    {"continent", type text},
                                    {"year", Int64.Type},
                                    {"lifeExp", type number},
                                    {"pop", Int64.Type},
                                    {"gdpPercap", type number}
                                                    }),
    filter = Table.SelectRows(col_types, each ([year] = 2007)),
    mutate = Table.AddColumn(filter, "lifeExpMonths",
                                  each [lifeExp] * 12, type number),
    arrange = Table.Sort(mutate,{{"lifeExpMonths", Order.Descending}})
in
```

R scripts

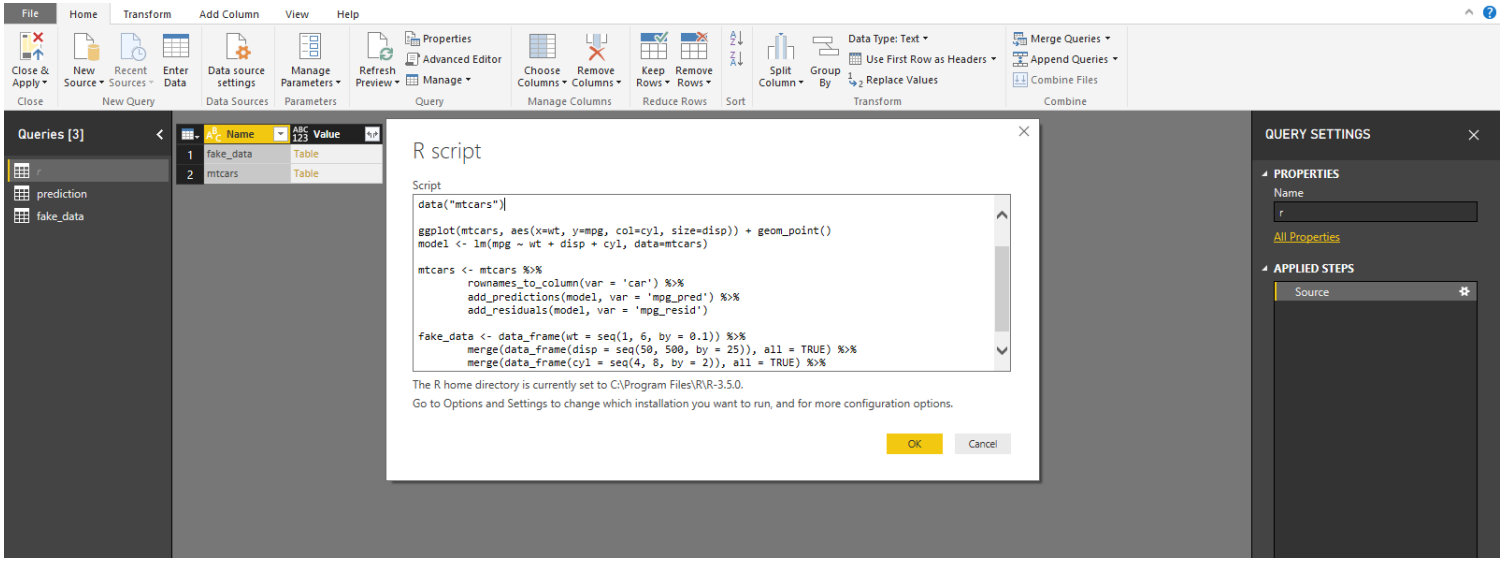
VS this

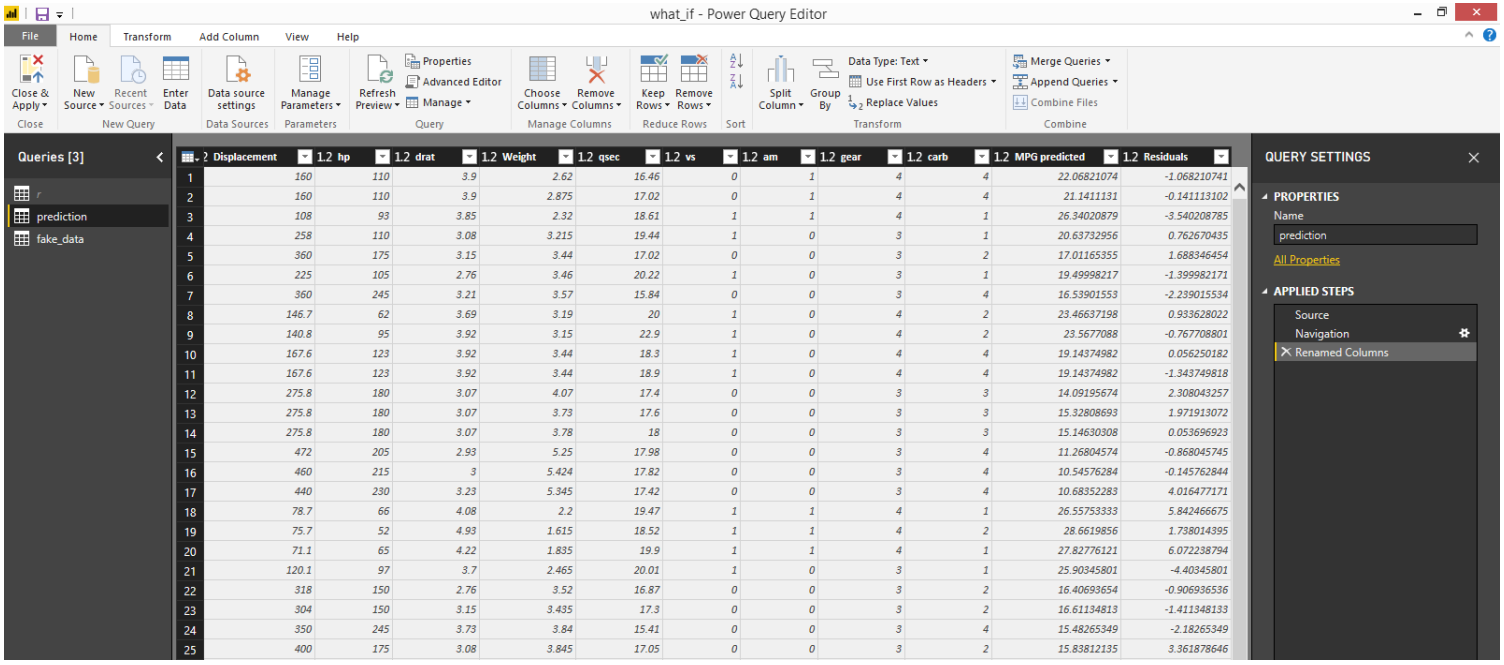
```
library(gapminder)
library(dplyr)

gapminder %>%
  filter(year == '2007') %>%
  mutate(lifeExpMonths = 12 * lifeExp) %>%
  arrange(desc(lifeExpMonths))
```

R scripts

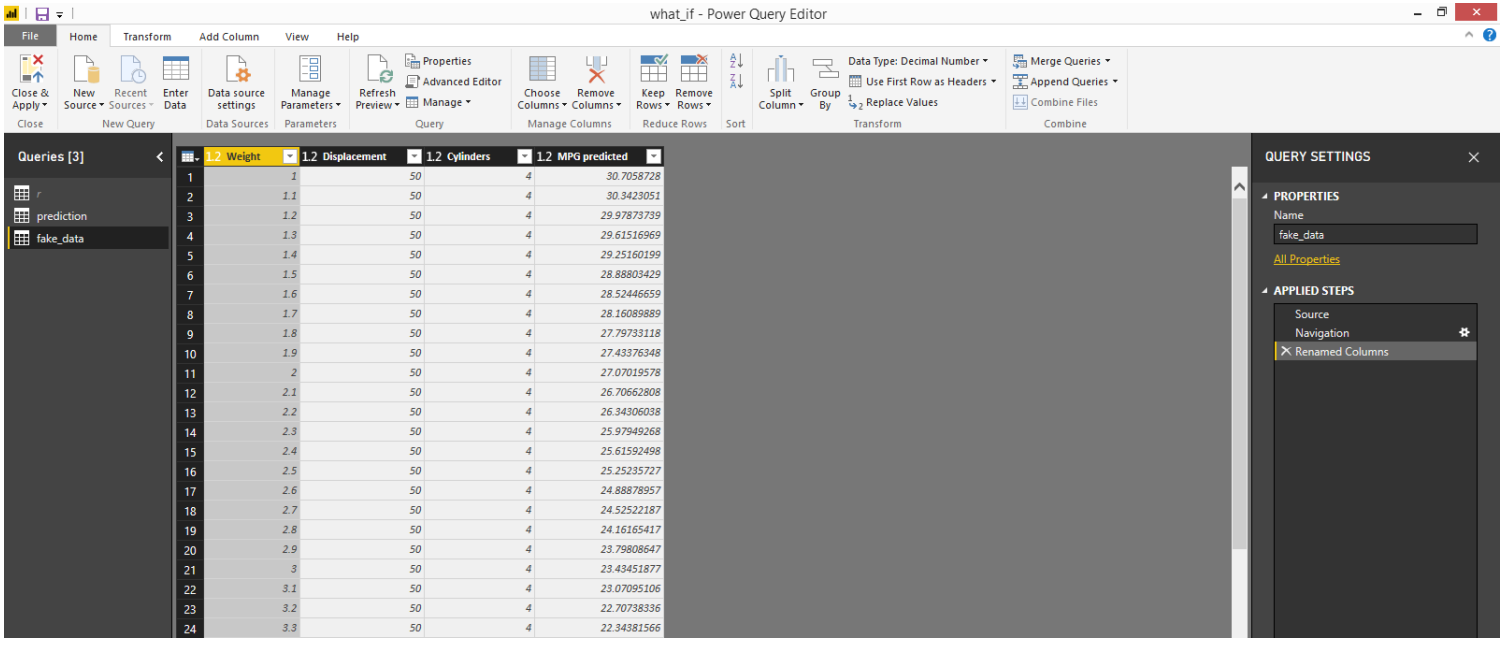
What If demo

- what_if.R
- what_if.pbix
- https://app.powerbi.com/view?r=eyJrIjoiNmNjZGE3MzEtMmMxYy00NDRmLWEyZmEtOTAyM2FmNTQzMzc2Iiwid


File
Home
Transform
Add Column
View
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Text
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [3]
r
prediction
fake_data
Name
Value
1 fake_data Table
2 mtcars Table
R script
Script
data("mtcars")
ggplot(mtcars, aes(x=wt, y=mpg, col=cyl, size=disp)) + geom_point()
model <- lm(mpg ~ wt + disp + cyl, data=mtcars)
mtcars <- mtcars %>%
rownames_to_column(var = 'car') %>%
add_predictions(model, var = 'mpg_pred') %>%
add_residuals(model, var = 'mpg_resid')
fake_data <- data_frame(wt = seq(1, 6, by = 0.1)) %>%
merge(data_frame(disp = seq(50, 500, by = 25)), all = TRUE) %>%
merge(data_frame(cyl = seq(4, 8, by = 2)), all = TRUE) %>%
The R home directory is currently set to C:\Program Files\R\R-3.5.0.
Go to Options and Settings to change which installation you want to run, and for more configuration options.
OK
Cancel
QUERY SETTINGS
PROPERTIES
Name
r
All Properties
APPLIED STEPS
Source



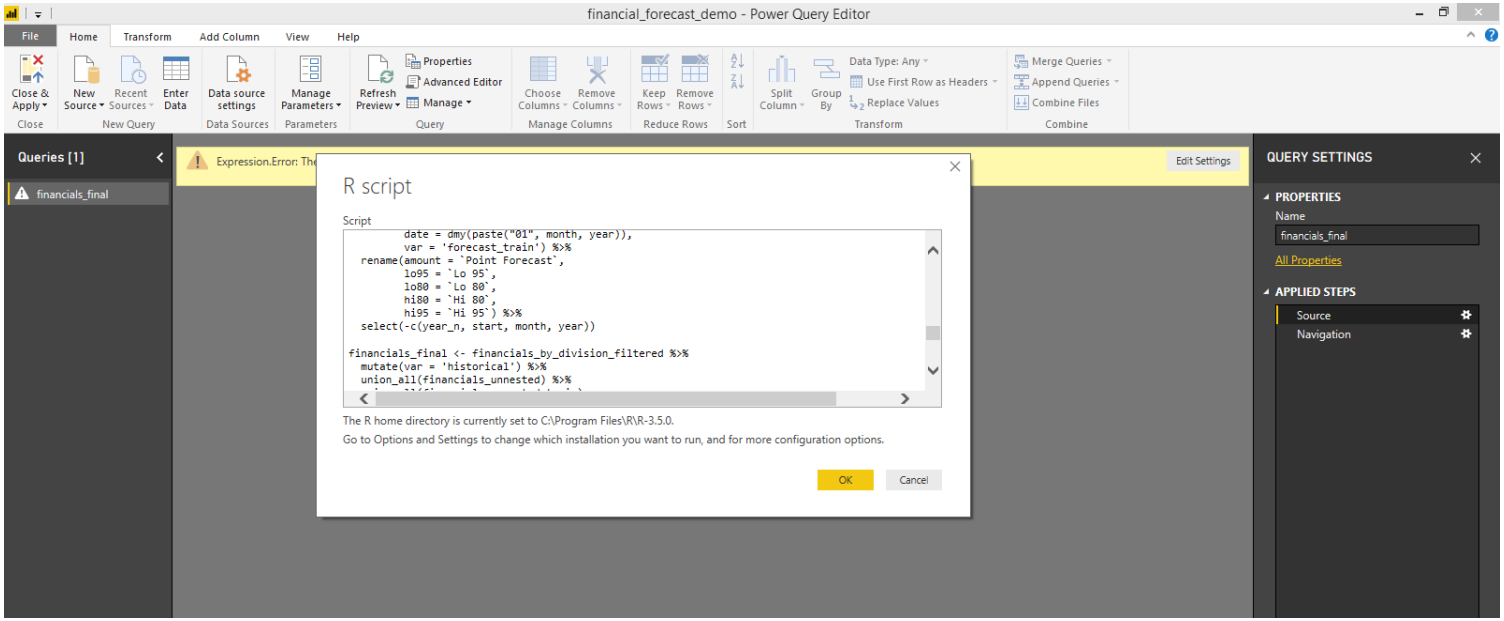
	Displacement	hp	drat	Weight	qsec	vs	am	gear	carb	MPG predicted	Residuals
1	160	110	3.9	2.62	16.46	0	1	4	4	22.06821074	-1.068210741
2	160	110	3.9	2.875	17.02	0	1	4	4	21.1411131	-0.141113102
3	108	93	3.85	2.32	18.61	1	1	4	1	26.34020879	-3.540208785
4	258	110	3.08	3.215	19.44	1	0	3	1	20.63732956	0.762670435
5	360	175	3.15	3.44	17.02	0	0	3	2	17.01165355	1.688346454
6	225	105	2.76	3.46	20.22	1	0	3	1	19.49998217	-1.399982171
7	360	245	3.21	3.57	15.84	0	0	3	4	16.53901553	-2.239015534
8	146.7	62	3.69	3.19	20	1	0	4	2	23.46637198	0.933628022
9	140.8	95	3.92	3.15	22.9	1	0	4	2	23.5677088	-0.767708801
10	167.6	123	3.92	3.44	18.3	1	0	4	4	19.14374982	0.056250182
11	167.6	123	3.92	3.44	18.9	1	0	4	4	19.14374982	-1.343749818
12	275.8	180	3.07	4.07	17.4	0	0	3	3	14.09195674	2.308043257
13	275.8	180	3.07	3.73	17.6	0	0	3	3	15.32808693	1.971913072
14	275.8	180	3.07	3.78	18	0	0	3	3	15.14630308	0.053696923
15	472	205	2.93	5.25	17.98	0	0	3	4	11.26804574	-0.868045745
16	460	215	3	5.424	17.82	0	0	3	4	10.54576284	-0.145762844
17	440	230	3.23	5.345	17.42	0	0	3	4	10.68352283	4.016477171
18	78.7	66	4.08	2.2	19.47	1	1	4	1	26.55753333	5.842466675
19	75.7	52	4.93	1.615	18.52	1	1	4	2	28.6619856	1.738014395
20	71.1	65	4.22	1.835	19.9	1	1	4	1	27.82776121	6.072238794
21	120.1	97	3.7	2.465	20.01	1	0	3	1	25.90345801	-4.40345801
22	318	150	2.76	3.52	16.87	0	0	3	2	16.40693654	-0.906936536
23	304	150	3.15	3.435	17.3	0	0	3	2	16.61134813	-1.411348133
24	350	245	3.73	3.84	15.41	0	0	3	4	15.48265349	-2.18265349
25	400	175	3.08	3.845	17.05	0	0	3	2	15.83812135	3.361878646

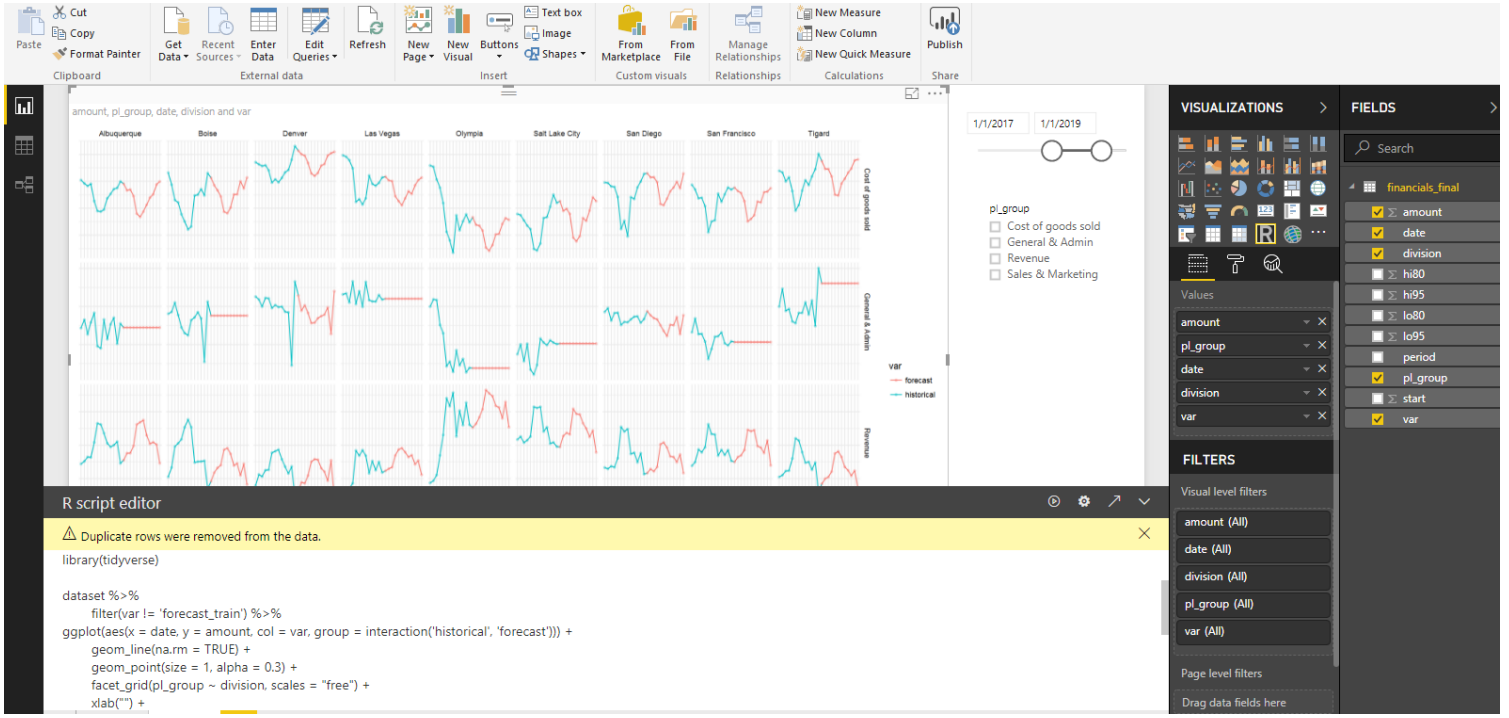

what_if - Power Query Editor
File
Home
Transform
Add Column
View
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Decimal Number
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [3]
r
prediction
fake_data
1.2 Weight
1.2 Displacement
1.2 Cylinders
1.2 MPG predicted
1 1 50 4 30.7058728
2 1.1 50 4 30.3423051
3 1.2 50 4 29.97873739
4 1.3 50 4 29.61516969
5 1.4 50 4 29.25160199
6 1.5 50 4 28.88803429
7 1.6 50 4 28.52446659
8 1.7 50 4 28.16089889
9 1.8 50 4 27.79733118
10 1.9 50 4 27.43376348
11 2 50 4 27.07019578
12 2.1 50 4 26.70662808
13 2.2 50 4 26.34306038
14 2.3 50 4 25.97949268
15 2.4 50 4 25.61592498
16 2.5 50 4 25.25235727
17 2.6 50 4 24.88878957
18 2.7 50 4 24.52522187
19 2.8 50 4 24.16165417
20 2.9 50 4 23.79808647
21 3 50 4 23.43451877
22 3.1 50 4 23.07095106
23 3.2 50 4 22.70738336
24 3.3 50 4 22.34381566
QUERY SETTINGS
PROPERTIES
Name
fake_data
All Properties
APPLIED STEPS
Source
Navigation
Renamed Columns

R scripts

Forecast demo

- financial_forecast_demo.pbix
- financial_forecast_demo.r
- https://app.powerbi.com/view?r=eyJrIjoiZWU0ZjUzZGUtZjk5Yi00ZGU1LThhZTgtYWJiNGEzYzFlZjgyIiwidCI6IjQ4NT


financial_forecast_demo - Power Query Editor
File
Home
Transform
Add Column
View
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Any
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [1]
financials_final
Expression.Error: Th
Edit Settings
R script
Script
date = dmy(paste("01", month, year)),
var = 'forecast_train') %>%
rename(amount = `Point Forecast`,
lo95 = `Lo 95`,
lo80 = `Lo 80`,
hi80 = `Hi 80`,
hi95 = `Hi 95`) %>%
select(-c(year_n, start, month, year))
financials_final <- financials_by_division_filtered %>%
mutate(var = 'historical') %>%
union_all(financials_unnested) %>%
The R home directory is currently set to C:\Program Files\R\R-3.5.0.
Go to Options and Settings to change which installation you want to run, and for more configuration options.
OK
Cancel
QUERY SETTINGS
PROPERTIES
Name
financials_final
All Properties
APPLIED STEPS
Source
Navigation


Paste
Cut
Copy
Format Painter
Clipboard
Get Data
Recent Sources
Enter Data
Edit Queries
Refresh
External data
New Page
New Visual
Buttons
Text box
Image
Shapes
Insert
From Marketplace
From File
Custom visuals
Manage Relationships
Relationships
New Measure
New Column
New Quick Measure
Calculations
Publish
Share
amount, pl_group, date, division and var
Albuquerque
Boise
Denver
Las Vegas
Olympia
Salt Lake City
San Diego
San Francisco
Tigard
Cost of goods sold
General & Admin
Revenue
var
forecast
historical
1/1/2017
1/1/2019
pl_group
Cost of goods sold
General & Admin
Revenue
Sales & Marketing
R script editor
Duplicate rows were removed from the data.
library(tidyverse)
dataset %>%
filter(var != 'forecast_train') %>%
ggplot(aes(x = date, y = amount, col = var, group = interaction('historical', 'forecast'))) +
geom_line(na.rm = TRUE) +
geom_point(size = 1, alpha = 0.3) +
facet_grid(pl_group ~ division, scales = "free") +
xlab("") +
VISUALIZATIONS
Values
amount
pl_group
date
division
var
FILTERS
Visual level filters
amount (All)
date (All)
division (All)
pl_group (All)
var (All)
Page level filters
Drag data fields here
FIELDS
Search
financials_final
amount
date
division
hi80
hi95
lo80
lo95
period
pl_group
start
var

Microsoft Power BI

1 of 2

When to use R in Power BI

Do

- performance
- machine learning, statistical analysis
- custom visualizations

Don't

- when no one else knows R on your team
- when you can get by with Power BI
- when you can get by with SQL

The end

Follow me:

- https://taraskaduk.com
- https://www.linkedin.com/in/taraskaduk/
- https://github.com/taraskaduk/
- https://twitter.com/taraskaduk

This presentation is here

https://github.com/taraskaduk/r-in-power-bi

https://taraskaduk.github.io/rpowerbi.html