**Analyzing Formula 1 results in R**

[Thomas Hütter](https://www.sqlservercentral.com/author/Thomas-Htter), 2020-03-27 (first published: 2018-08-21)

FROM: <https://www.sqlservercentral.com/articles/analyzing-formula-1-results-in-r>

This is my first article on SQLServerCentral (if you like it, there'll probably be more...). I am a software developer, I deal with SQL Server databases regularly and I've been playing around with R for some time. I even do presentations at SQLSaturdays and other events, which caught [Steve Jones](http://www.sqlservercentral.com/Authors/Articles/Steve_Jones/3/)' attention, and he asked me if I was interested in writing some articles for SSC. Well, yes, so here we go.

This will not be the absolute beginner's first lesson, so if you wanna play along, I trust you have R and [RStudio](https://www.rstudio.com/) installed and know what the assignment operator looks like in R. Today's task is to fetch some data from the internet, bring it into shape, und visualize it. As I have no interest whatsoever in football, I went for the next-obvious sport for my data source: the Formula 1 racing series. I attached the R script I use, which hopefully contains the right dose of comments.

First we load a couple of packages.

# Preload needed packages -------------------------------------------------

library(rvest) # For scraping the web

library(tidyr) # Functions to tidy our data

library(dplyr) # Pipe operator, tibble handling

library(ggplot2) # The Swiss army knife of plotting in R

Remember, if any of the packages is missing on your system, you can install them one-by-one or get the tidyverse family (which I'll probably explain in an upcoming article) in one go:

install.packages("tidyverse")

Next up, take a look at the internet site we will get the data from. There is an R function to open it in your browser :-). We find all kinds of distractions, such as buttons, menus and advertisements. Luckily the race result data is formatted as HTML table, so we can easily use some R functions to get the whole HTML construct, isolate the table element, extract just the raw data, and display it.

# Import/ingest the Formula 1 race results for season 2016 ----------------

# Take a look at the data in the browser

browseURL('https://www.formel1.de/saison/wm-stand/2016/fahrer-wertung')

# Fetch the contents of the HTML-table into the variable f1

f1 <- read\_html('https://www.formel1.de/saison/wm-stand/2016/fahrer-wertung') %>%

html\_node('table') %>%

html\_table()

# Display our data

f1

So our code has rather nicely separated the data we wanted from those colourful distractions, but it's far from being perfect. The result columns come without any headers, and there are minus signs where we would prefer having zeros. We'll fix that and, for reasons you'll see shortly, we'll cut the results down to the best 9 drivers. Also, for our plot functions to work smoothly, we convert the whole thing to the "long" format.

# Transform & tidy the data -----------------------------------------------

# Add missing column headers

colnames(f1) <- c('Pos', 'Driver', 'Total', sprintf('R%02d', 1:21))

# Convert to tibble data frame and filter on top 9 drivers

f1 <- as\_tibble(f1) %>%

filter(as.integer(Pos) <= 9)

# Make Driver a factorial variable, replace all '-' with zeros, convert to long format

f1$Driver <- as.factor(f1$Driver)

f1[, -2] <- apply(f1[, -2], 2, function(x) as.integer(gsub('-', '0', as.character(x))))

f1long <- gather(f1, Race, Points, R01:R21)

# That looks better

f1long

Now we are ready for some plots, first the "big" picture. The ggplot function is relatively straightforward to use: first parameter is the data frame we want to use. The aethetics basically determine which variables go on the x- and y-axis, optionally which variable to group on and which to colour by. Then we have to state which geometry to plot, in our case a line plot. The rest is some cosmetics, as uncluttering the ticks on the x axis and giving the plot a title.

# Now for some plots ------------------------------------------------------

# Everything in the 'big picture'

ggplot(f1long, aes(x = Race, y = Points, group = Driver, colour = Driver)) +

geom\_line() +

scale\_x\_discrete(breaks=c('R01', 'R06', 'R11', 'R16', 'R21')) +

labs(title = 'F1 race results 2016, top 9 drivers',

caption = 'source: www.formel1.de')

Here we have the 2016 season's results in one plot. Despite having limited our data to the 9 best drivers, it is, well, sort of cluttered. Can we do anything about that? Sure we can... possible solution coming up, something called facetting:

The second plot actually contains the exact same information as the previous one. Just - every driver gets his own small sub-plot, so you can easily follow how each of them performed during the season. And thanks to the auxiliary lines (the white lines originating from each axis tick), the results can still easily be compared to each other.

# Each drivers gets his own

ggplot(f1long, aes(x = Race, y = Points, group = Driver, colour = Driver)) +

geom\_line(show.legend = FALSE) + facet\_wrap(~ Driver) +

scale\_x\_discrete(breaks=c('R01', 'R06', 'R11', 'R16', 'R21')) +

labs(title = 'F1 race results 2016, top 9 drivers',

caption = 'source: www.formel1.de')

The plots show how each driver performed, that is how many points they gained in the races from 1 through 21. You can see that Ricciardo and Verstappen won one race each, the ones where they gained 25 points. The real top-scorers were Hamilton and Rosberg though, with Rosberg having collected enough points after race 17, he could let Hamilton win the remaining races and still end the season as champion.

I hope I could give you a first impression of what is possible in R with relatively little coding effort. Especially if you look at the ggplot calls: once our data was in shape, those were basically two-liners plus a little cosmetic. We saw that even a data set with just 9 'groups' can be too much for one plot. But then, the step of decluttering the plot consisted only of adding a call to facet\_wrap and switching off the legend.

That's it for today. If you enjoyed our little project, please leave a comment. If you didn't like it - please also leave a comment and tell me where I can improve. Thanks for now!

**Resources**

* [formula1.r](https://www.sqlservercentral.com/wp-content/uploads/2019/05/formula1.r)
* about the item [Analyzing Formula 1 results in R](https://www.sqlservercentral.com/articles/analyzing-formula-1-results-in-r)
* [Phil Parkin](https://www.sqlservercentral.com/forums/user/phil-parkin)

SSC Guru

Points: 244105

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/124104)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002651&_wpnonce=9ed9221412#new-post)

August 21, 2018 at 11:00 am

[#2002651](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002651)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002651&_wpnonce=9ed9221412#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002651&_wpnonce=6cad59f228) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002651)

**Thomas HÃ¼tter - Monday, August 20, 2018 11:28 PM**

Comments posted to this topic are about the item [**Analyzing Formula 1 results in R**](https://www.sqlservercentral.com/articles/R+Language/175012/)

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

If the answer to your question can be found with a brief Google search, please perform the search yourself, rather than expecting one of the SSC members to do it for you.

* [jonathan.crawford](https://www.sqlservercentral.com/forums/user/jonathancrawford)

SSCertifiable

Points: 6516

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/495870)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002670&_wpnonce=26e4da9d54#new-post)

August 21, 2018 at 12:39 pm

[#2002670](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002670)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002670&_wpnonce=26e4da9d54#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002670&_wpnonce=878764339e) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002670)

**Phil Parkin - Tuesday, August 21, 2018 6:00 AM**

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

I'm out of practice Phil, but I'm sure it can, you used to be able to use an ODBC connection to do just that. the package I was using for that went bust and I haven't found a new one that works for me with my version yet, but I don't play with it very much.

-------------------------------------------------------------------------------------------------------------------------------------  
Please follow [Best Practices For Posting On Forums](http://www.sqlservercentral.com/articles/Best+Practices/61537/) to receive quicker and higher quality responses

* [Phil Parkin](https://www.sqlservercentral.com/forums/user/phil-parkin)

SSC Guru

Points: 244105

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/124104)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002674&_wpnonce=847d124ddd#new-post)

August 21, 2018 at 12:53 pm

[#2002674](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002674)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002674&_wpnonce=847d124ddd#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002674&_wpnonce=1327f4b01d) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002674)

**jonathan.crawford - Tuesday, August 21, 2018 7:39 AM**

**Phil Parkin - Tuesday, August 21, 2018 6:00 AM**

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

I'm out of practice Phil, but I'm sure it can, you used to be able to use an ODBC connection to do just that. the package I was using for that went bust and I haven't found a new one that works for me with my version yet, but I don't play with it very much.

As far as I know, the ODBC connection from R Studio to SQL Server facilitates the import of SQL Server data into R Studio prior to running R scripts against that data.

But my knowledge is just at beginner level and I may be wrong. If I am wrong, and R Studio can query SQL Server data directly, it almost makes redundant the work which MS did to integrate R into SQL Server.

If the answer to your question can be found with a brief Google search, please perform the search yourself, rather than expecting one of the SSC members to do it for you.

* [Thomas Hütter](https://www.sqlservercentral.com/forums/user/Thomas-Htter)

SSC Eights!

Points: 849

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/369772)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002679&_wpnonce=e0898e21bd#new-post)

August 21, 2018 at 1:16 pm

[#2002679](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002679)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002679&_wpnonce=e0898e21bd#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002679&_wpnonce=535beb6bfe) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002679)

**Phil Parkin - Tuesday, August 21, 2018 6:00 AM**

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

Thanks Phil, and I apologize for raising false hopes there. I use RStudio for 99% of my R stuff, for three reasons:

* + the use of "sp\_execute\_external\_script" in SQL Server Management Studio is a bit clumsy for my taste,
  + I don't use R functionality in any production environment so far, so I haven't had the need to handle "big" data,
  + for private purposes and for sidekick projects, I use a Mac (which doesn't natively run the full-featured SQL Server).

Do I understand you correctly: you would like to see my example code run inside SSMS, including the plot output?

* [rchantler](https://www.sqlservercentral.com/forums/user/rchantler)

SSCrazy

Points: 2116

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/520809)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002681&_wpnonce=69bd5e889a#new-post)

August 21, 2018 at 1:22 pm

[#2002681](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002681)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002681&_wpnonce=69bd5e889a#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002681&_wpnonce=7a499ec039) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002681)

Hi Thomas,

Very nice article, thank you.  
I just wanted to add my voice to the others who wanted to see all of this done from within SSMS, i.e. making use of the R integration.

* [Phil Parkin](https://www.sqlservercentral.com/forums/user/phil-parkin)

SSC Guru

Points: 244105

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/124104)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002683&_wpnonce=fd2c87df23#new-post)

August 21, 2018 at 1:26 pm

[#2002683](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002683)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002683&_wpnonce=fd2c87df23#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002683&_wpnonce=5f71757d36) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002683)

**Thomas HÃ¼tter - Tuesday, August 21, 2018 8:16 AM**

**Phil Parkin - Tuesday, August 21, 2018 6:00 AM**

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

Thanks Phil, and I apologize for raising false hopes there. I use RStudio for 99% of my R stuff, for three reasons:

* + the use of "sp\_execute\_external\_script" in SQL Server Management Studio is a bit clumsy for my taste,
  + I don't use R functionality in any production environment so far, so I haven't had the need to handle "big" data,
  + for private purposes and for sidekick projects, I use a Mac (which doesn't natively run the full-featured SQL Server).

Do I understand you correctly: you would like to see my example code run inside SSMS, including the plot output?

It would be the icing on the cake for your article, yes. But I do not think that it can be done, based on my current rudimentary knowledge.

If there is any way at all of being able to run R scripts against data stored natively in SQL Server & then plotting the results, that would be a huge win. Without importing, of course.

If the answer to your question can be found with a brief Google search, please perform the search yourself, rather than expecting one of the SSC members to do it for you.

* [Thomas Hütter](https://www.sqlservercentral.com/forums/user/Thomas-Htter)

SSC Eights!

Points: 849

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/369772)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002684&_wpnonce=f714fdf723#new-post)

August 21, 2018 at 1:30 pm

[#2002684](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002684)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002684&_wpnonce=f714fdf723#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002684&_wpnonce=844b843c38) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002684)

OK folks, please take into account that I am in a different time zone, and have to finish my day work first.   
I'll get back to you with a SQL script (or probably two) in the evening (European time...).

* [rchantler](https://www.sqlservercentral.com/forums/user/rchantler)

SSCrazy

Points: 2116

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/520809)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002685&_wpnonce=daed596fc6#new-post)

August 21, 2018 at 1:49 pm

[#2002685](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002685)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002685&_wpnonce=daed596fc6#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002685&_wpnonce=a11776de19) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002685)

Maybe a second article?

* [Phil Parkin](https://www.sqlservercentral.com/forums/user/phil-parkin)

SSC Guru

Points: 244105

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/124104)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002687&_wpnonce=27bb4dbe65#new-post)

August 21, 2018 at 1:50 pm

[#2002687](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002687)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002687&_wpnonce=27bb4dbe65#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002687&_wpnonce=cc96a79a12) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002687)

**Thomas HÃ¼tter - Tuesday, August 21, 2018 8:30 AM**

OK folks, please take into account that I am in a different time zone, and have to finish my day work first.   
I'll get back to you with a SQL script (or probably two) in the evening (European time...).

Thomas, please don't feel the need to rush things ... I did not intend to appear at all demanding.

If the answer to your question can be found with a brief Google search, please perform the search yourself, rather than expecting one of the SSC members to do it for you.

* [Thomas Hütter](https://www.sqlservercentral.com/forums/user/Thomas-Htter)

SSC Eights!

Points: 849

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/369772)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002762&_wpnonce=1be2b6a85e#new-post)

August 21, 2018 at 7:30 pm

[#2002762](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002762)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002762&_wpnonce=1be2b6a85e#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002762&_wpnonce=099bc87b5e) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002762)

Ha ha, no worries - I was prepared to answer some questions today. And rest assured, I enjoy what I am doing here.

OK, so to make my example work in SSMS, and answer Phil's questions, I produced two (+ one) SQL scripts. I had to rename them to .txt to upload them here, so you'll probably have to rename them back to the .sql extension before using them. And there are some caveats: you have to install the needed R packages into your SQL Server R service first (use my installpkg script). But in order to accomplish that, you have to have write permissions to the library folder, in my case of a standard installation, that is "C:\Program Files\Microsoft SQL Server\MSSQL14.MSSQLSERVER\R\_SERVICES\library". I'll leave it up to you to sort that out.

Then, you can run f1getdata in SSMS. If all is well, this will produce a SQL table called f1table, that holds the brushed up long version of the formula 1 data. Check if you have 189 records in there. The INSERT INTO statement on the SQL side takes the output of the "EXEC sp\_execute\_external\_script" call. That is the way to persist data from your R script into SQL Server. (As I ran into some character encoding issues in the second script, I implemented a workaround converting umlauts - that's the two UPDATE statements at the bottom of the first script. )

The f1plotdata script needed some additions to the R code. This is because I do not know a feasible, easy way to show R plots within SSMS. I prepare the two plots to be saved in separate .png files in the Temp directory. I hope it is easy to spot the principle how to embed a SELECT statement into a variable and then use it as input to the R script by assigning it to the @input\_data\_1 parameter. After running this script, you should find Plot1.png and Plot2.png in your C:\Temp folder.

I hope I could shed some light on how R scripts work in SSMS...?

* [xsevensinzx](https://www.sqlservercentral.com/forums/user/xsevensinzx)

One Orange Chip

Points: 25551

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/1672113)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002775&_wpnonce=7ebcaf1526#new-post)

August 21, 2018 at 10:33 pm

[#2002775](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002775)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002775&_wpnonce=7ebcaf1526#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002775&_wpnonce=c4d6b79b40) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002775)

**Phil Parkin - Tuesday, August 21, 2018 8:26 AM**

**Thomas HÃ¼tter - Tuesday, August 21, 2018 8:16 AM**

**Phil Parkin - Tuesday, August 21, 2018 6:00 AM**

Nice article, Thomas.  
I must admit that I was expecting/hoping to see something different, though.  
I was hoping that you would   
a) Import the data into a SQL Server table  
b) Use sp\_execute\_external\_script to run the various R scripts against the imported data.

I'd be particularly interested to know whether the plotting options available in R Studio can somehow be used on data held in SQL Server, without the need for moving the data out of SQL Server.

Thanks Phil, and I apologize for raising false hopes there. I use RStudio for 99% of my R stuff, for three reasons:

* + the use of "sp\_execute\_external\_script" in SQL Server Management Studio is a bit clumsy for my taste,
  + I don't use R functionality in any production environment so far, so I haven't had the need to handle "big" data,
  + for private purposes and for sidekick projects, I use a Mac (which doesn't natively run the full-featured SQL Server).

Do I understand you correctly: you would like to see my example code run inside SSMS, including the plot output?

It would be the icing on the cake for your article, yes. But I do not think that it can be done, based on my current rudimentary knowledge.

If there is any way at all of being able to run R scripts against data stored natively in SQL Server & then plotting the results, that would be a huge win. Without importing, of course.

Ideally, what you would do is run the R scripts on the data in SQL Server in order to extract the coefficients (the model) to a physical table. Then you can combine this with the data in SQL Server via a view and plot the data directly. Far as I know, Azure DB nor say, Azure ML has this ability.

And just for clarity, you don't just do this because it's "Big Data". You do this so you can make dynamic forecasting more fluid on the source data as well allow you to repeat the process in a more consistent fashion (i.e.: move from ad-hoc R analysis to enterprise).

You also would move away from ggplot and move more into something like Power BI or Tableau. RStudio is only good for script/adhoc analysis.

* [Thomas Hütter](https://www.sqlservercentral.com/forums/user/Thomas-Htter)

SSC Eights!

Points: 849

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/369772)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002795&_wpnonce=e226897c8f#new-post)

August 22, 2018 at 5:53 am

[#2002795](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002795)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002795&_wpnonce=e226897c8f#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002795&_wpnonce=228ac75c0a) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002795)

**xsevensinzx - Tuesday, August 21, 2018 5:33 PM**

RStudio is only good for script/adhoc analysis.

Thanks for confirming - that's exactly what my article is about: giving a small, yet reproducible example for data analysis using R.  
No model, no predictions, no data that would be too big for my MacBook's RAM.

* [Steve Jones - SSC Editor](https://www.sqlservercentral.com/forums/user/steve-jones-ssc-editor)

SSC Guru

Points: 717788

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/51)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002871&_wpnonce=4216a003cc#new-post)

August 22, 2018 at 3:27 pm

[#2002871](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002871)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002871&_wpnonce=4216a003cc#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002871&_wpnonce=4b16d56b0c) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002871)

Be good to see a second article doing this in SQL Server

If any of you would like to see it,  encourage Thomas to write more.

Follow me on Twitter: [http://www.twitter.com/way0utwest](https://www.twitter.com/way0utwest)  
[Forum Etiquette: How to post data/code on a forum to get the best help](https://www.sqlservercentral.com/articles/forum-etiquette-how-to-post-datacode-on-a-forum-to-get-the-best-help)  
My Blog: [www.voiceofthedba.com](http://www.voiceofthedba.com/)

* [kenstevenson](https://www.sqlservercentral.com/forums/user/kenstevenson)

SSC Enthusiast

Points: 160

[More actions](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Follow](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#follow)

[Send a Message](https://www.sqlservercentral.com/forums/user/WJBergen/messages/w/19902)

[Block user](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002893&_wpnonce=5141ae47c1#new-post)

August 22, 2018 at 5:10 pm

[#2002893](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r/#post-2002893)

[Reply](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?bbp_reply_to=2002893&_wpnonce=5141ae47c1#new-post) | [Report](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r?action=bbp_rc_toggle_reply_report&reply_id=2002893&_wpnonce=276d724081) | [Quote](https://www.sqlservercentral.com/forums/topic/analyzing-formula-1-results-in-r#2002893)

Good work! I am quite an enthusiast of both F1 and SQL Server. Guess which one pays the bills. My company does use R, yet I have not had the opportunity to work myself into an R project. Looking forward to future articles.

I happened to be quite pleased that the topic was in fact based around Formula 1. Then, a bit disappointed that you did not somehow uncover share a mountain of F1 telemetry data