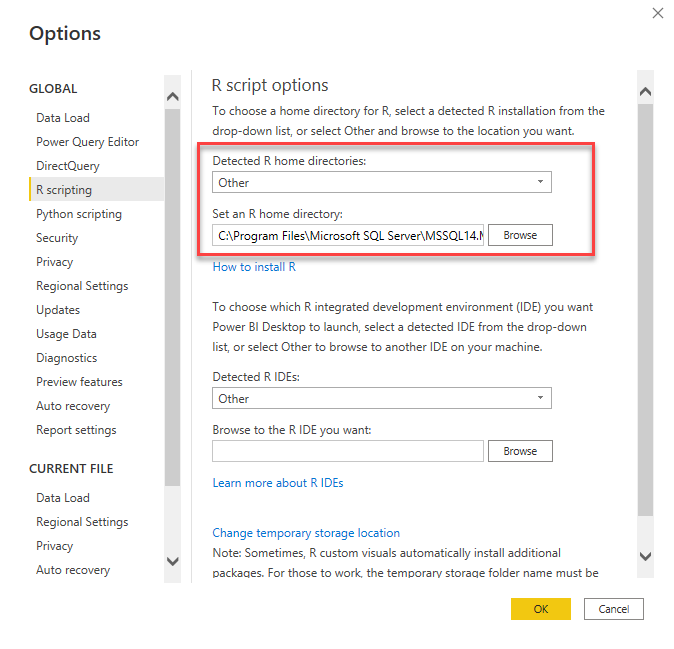
# My environment

## Power BI Desktop

* Download from <https://www.microsoft.com/en-us/download/details.aspx?id=58494>
* Open the Customer Analysis.pbix file. You should be able to use the file without connecting to SQL Server as long as you don’t try to refresh the file or execute the R script.
  + To update the file to connect to your own SQL Server instance:
    - Click Transform Data on the Home tab of the ribbon.
    - Click Data Source Settings on the Home tab of the ribbon in Power Query Editor.
    - With AdventureWorksDW selected, click the Change Source button.
    - Change the server and database settings to match your environment if you have a SQL Server with AdventureWorks DW installed (see below).
* To refresh the R visuals, you will need an installation of R on your computer and you need to configure the home directory by opening Options and Settings from the File menu, selecting Options, and then R Scripting. Configure the home directory by selecting it in the Detected R Home Directories drop-down list, or keep Other selected and manually set the path to the directory.



* If you open the demonstration PBIX file, you will be prompted if you want to enable scripts. This is due to the presence of R visuals in the file.

Graphical user interface, text

Description automatically generated

## SQL Server 2019 with R Services

* Download from <https://www.microsoft.com/en-us/download/details.aspx?id=100809>
* Install R Services as described here: <https://docs.microsoft.com/en-us/sql/machine-learning/install/sql-machine-learning-services-windows-install?view=sql-server-ver15>

## SQL Server Management Studio

* Download from [Download SQL Server Management Studio (SSMS) - SQL Server Management Studio (SSMS) | Microsoft Docs](https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?redirectedfrom=MSDN&view=sql-server-ver15)
* Refer to instructions here to install AdventureWorksDW from a backup. A link to the backup file is on this page ([sql-server-samples/samples/databases/adventure-works at master · microsoft/sql-server-samples · GitHub](https://github.com/microsoft/sql-server-samples/tree/master/samples/databases/adventure-works)) in addition to “restore a database backup” instructions. Be sure to get the AdventureWorksDW20xx.bak file. I believe I used the 2014 version, but any version should be fine. The main difference would be in the date range used for sales.

# Code Snippets

## Age histogram – version 1

hist(dataset$Age)

## YearlyIncome histogram – version 1

hist(dataset$YearlyIncome)

## Sales histogram – version 1

hist(dataset$SalesAmount)

## Age histogram – version 2

layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))

par(mar=c(0, 3.1, 1.1, 2.1))

boxplot(dataset$Age, horizontal=TRUE , xaxt="n" , col=rgb(0.8,0.8,0,0.5) , frame=F)

par(mar=c(4, 3.1, 1.1, 2.1))

hist(dataset$Age, col=rgb(0.2,0.8,0.5,0.5) )

# Add median to histogram

abline(v = median(dataset$Age), col = "blue", lwd = 3)

abline(v = mean(dataset$Age), col = "red", lwd = 3)

# Add text to histogram

text(x = 80, y = 2000, paste("Median of age=", median(dataset$Age)), col = "blue", cex=2)

text(x = 80, y = 2500, paste("Mean of age=", mean(dataset$Age)), col = "red", cex=2)

## YearlyIncome histogram – version 2

layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))

par(mar=c(0, 3.1, 1.1, 2.1))

boxplot(dataset$YearlyIncome, horizontal=TRUE , xaxt="n" , col=rgb(0.8,0.8,0,0.5) , frame=F)

par(mar=c(4, 3.1, 1.1, 2.1))

hist(dataset$YearlyIncome, col=rgb(0.2,0.8,0.5,0.5) )

# Add median to histogram

abline(v = median(dataset$YearlyIncome), col = "blue", lwd = 3)

abline(v = mean(dataset$YearlyIncome), col = "red", lwd = 3)

# Add text to histogram

text(x = 120000, y = 1500, paste("Median of yearly income=", median(dataset$YearlyIncome)), col = "blue", cex=2)

text(x = 120000, y = 1000, paste("Mean of yearly income=", mean(dataset$YearlyIncome)), col = "red", cex=2)

## Sales histogram – version 2

layout(mat = matrix(c(1,2),2,1, byrow=TRUE), height = c(1,8))

par(mar=c(0, 3.1, 1.1, 2.1))

boxplot(dataset$SalesAmount, horizontal=TRUE , xaxt="n" , col=rgb(0.8,0.8,0,0.5) , frame=F)

par(mar=c(4, 3.1, 1.1, 2.1))

hist(dataset$SalesAmount, col=rgb(0.2,0.8,0.5,0.5) )

# Add median to histogram

abline(v = median(dataset$SalesAmount), col = "blue", lwd = 3)

abline(v = mean(dataset$SalesAmount), col = "red", lwd = 3)

# Add text to histogram

text(x = 6000, y = 4000, paste("Median of sales=", median(dataset$SalesAmount)), col = "blue", cex=2)

text(x = 6000, y = 3000, paste("Mean of sales=", mean(dataset$SalesAmount)), col = "red", cex=2)