MARKET *Mix Modeling*

Digital Cognition(R)
(https://github.com/digital-cognition-co-in)

TV , Website,
Twitter
Facebook ...
Placement and
Promotion

Multi Channel Ad Spend

Available channels for Ad Spend …

Digital Cognition(R)
(https://github.com/digital-cognition-co-in)

Market ..

the action or business of promoting our products or services ...

Mix ..

..the types of promotion techniques and medium. TV , Social Media , YouTube , Newspapers and Mags...

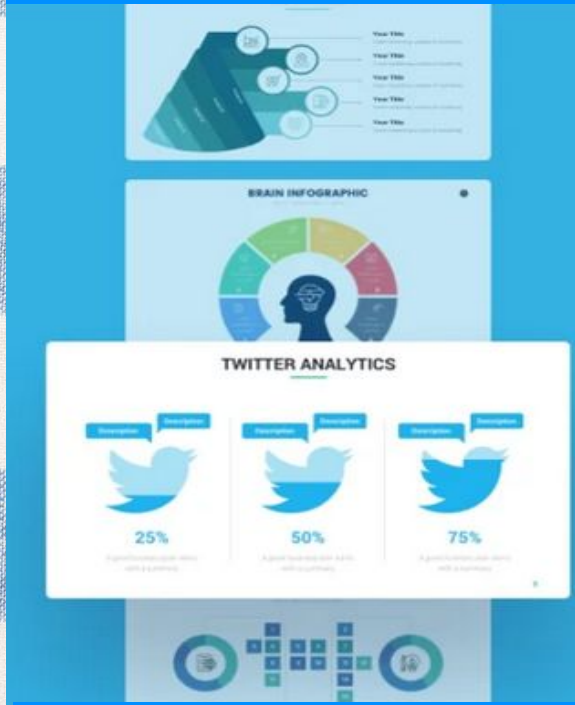
Modeling ..

relationships between various types of information (‘the mix part’) and inferring results .

The Statistical Tools

Why R and Python ?

The statistical tools of Choice - R and Python

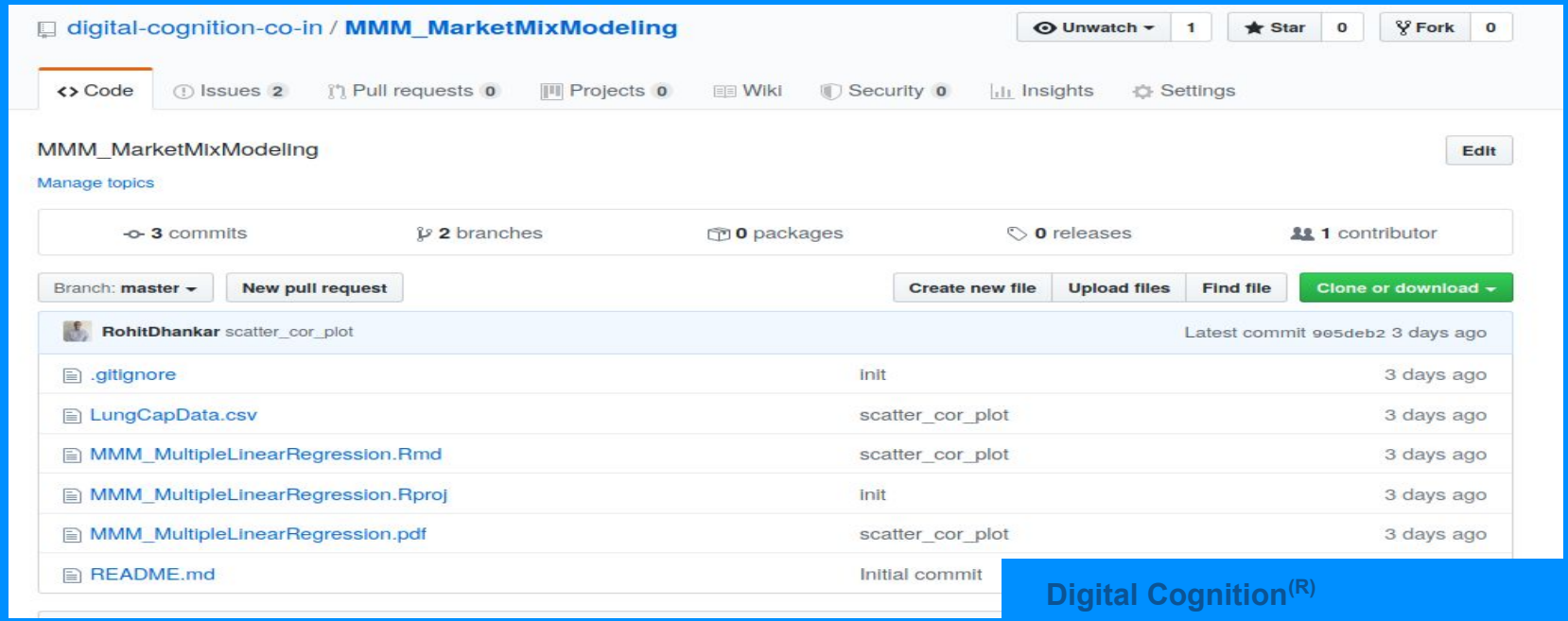


Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

R & Python Code

https://github.com/digital-cognition-co-in



Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

The Conceptual Framework

Why MMM ?

The Conceptual background for Market Mix Modeling

Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

Linear Relationship - Dependent (Sales) & Independent Variables.

Failure to Isolate impact of Independent variables .

Non Linear Relationship - Decay of impact.

Linear Relation

Isolate Impact | ROI ambiguity | Non Linear Relation

Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

MMM... as a statistical inference can simply be a case of **Multiple Linear Regression** or even advanced algorithms**...**

Estimate impact of various marketing tactics (marketing mix) on sales and predict / forecast - impact of future sets of tactics.

Used to ***optimize advertising mix*** and ***marketing budget spends***

Digital Cognition(R)
(https://github.com/digital-cognition-co-in)

ADSTOCK

Ad Build and Decay

Adstock

Gross Rating Point (GRP)

Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

Multiple Linear Regression

Sales (Dependent or Response Var.)

$$Sales_t = B_0 + B_1 Road..Hoard..Imp_t + B_2 TvAdSpend_t + B_p x_p + e_t$$

$Sales_t$ = at given time-**t**,is Linear Function of right side of **Multiple Linear Regression Model** shown above.

Parameters $(B_0, B_1, B_2 \ldots B_p)$ B_1 onwards for $p(x_p)$ independent variables.
epsilon (e_t) the Error or Random term -the catch all(measure errors + unknown var's),accounts for all kinds of variability in **$Sales_t$** which can't, be explained,by linear effect of -*p*-indep var's.

- B_0 = Intercept on Y Axis , is expected value of Y when all other Ind var == 0
- B_1 = Param for RoadsideHoardingsImp..(Independent Var.) at time - t. It represents ESTIMATE of Change in $Sales_t$ (Y-Dependent var.) corresponding to One Unit change in quant measure of - RoadsideHoardingsImpressions , when ***all other Independent Variables are held constant*** or are considered to be Constant.
- B_2 = Param for the TvAdSpend (Independent Var.)

Multiple Linear Regression

Sales (Dependent or Response Var.)

$$Sales_t = B_0 + B_1 Road..Hoard..Imp_t + B_2 TvAdSpend_t + B_p x_p + e_t$$

Multiple Linear Regression Model (shown above.)

$$E(y) = B_0 + B_1 x_1 + B_2 x_2 + \ldots + B_p x_p$$

Multiple Linear Regression Equation -

Assumed that ***mean or expected value*** of ***e (Epsilon)*** = 0

$$y\text{-}hat = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$$

Estimated Multiple Linear Regression Equation -

where ($b_1 + b_2 + \ldots + b_p$) - are estimates of the population Params as population Params are usually not known .

Why - Multiple Linear Regression ?

Sales (Dependent or Response Var.)

How is a **Multiple Linear Regression Model** a better predictor of the Sales (Dependent or Response Var.) than maybe a **Simple Linear Regression Model** ?

The SSR and SSE Values will change when we add - Multiple Independent variables , thus we infer that a **Multiple Linear Regression Model** is a better predictor of the Sales.

SST = Total Sum of Squares

Relation between the SST, SSR and SSE :-

$$SST = SSR + SSE$$

$$SST = \text{total sum of squares} = \Sigma(y_i - ybar)^2$$

$$SSR = \text{sum of squares due to regression} = \Sigma(yhat_i - ybar)^2$$

$$SSE = \text{sum of squares due to error} = \Sigma(y_i - yhat_i)^2$$

R-squared - *Multiple* Coefficient Of Determination

Multiple **Coefficient Of Determination (R^2 R-squared)**

$$R^2 = SSR/SST$$

Proportion of variability in the **dependent variable** which is **explained** by the **Estimated multiple regression equation.** Its generally observed that **R-squared will increase as we add variables to a - multiple regression equation.**

Thus for MMM we will get a Higher value for R^2 with - $y \sim RoadsideHoardingsImp + TvAdSpend + x_p$

Vs.

$y \sim RoadsideHoardingsImp$ ….only

Adjusted-R-squared

Adjusted Multiple **Coefficient Of Determination ($R^2_{(adj)}$ Adjusted-R-squared)**

$$R^2_{adj} = 1-(1-R^2)n-1/n-p-1$$

n = count of Obs.
p = count of independent var.

To compensate for the addition of Multiple var's **R-squared to** the model we need to factor in the Adjusted **R-squared**

F-test Test for Overall Significance

Within the scope of **Multiple Linear regression - the F Test** is called the **- Test for Overall Significance ,** as it determines if a **Significant Relationship** exists between the **Dependent Var and All the Independent Var's .** After the F-test is used to **- Test for Overall Significance ,** a separate **t-test** is conducted for all independent variables.

$H_0 = B_0 = B_1 = B_2 = B_p = 0$
This Null Hypothesis - states that all PARAMS are Equal to Zero

H_a
The alternative Hypothesis - states that at least one PARAM is Not Equal to Zero
We want to Reject the Null Hypothesis and accept the alternative.

This leads us to **MEAN SQUARE ERRORS** - the **MSE** and the **MSR**

F-test Test for Overall Significance

MEAN SQUARE ERRORS - the **MSE (Mean Square due to Error)** and the **MSR (Mean Square due to Regression)**

$$MSR = SSR / p$$

$$MSE = SSE / n-p-1$$

$$F = MSR / MSE$$

Total degrees of Freedom = n - 1 (n = sample size)
Degrees of Freedom - Residual Error = (n-1)-(count of Ind Var)

Analysis of variance - F-test (Not in scope here)

t-test Test for Individual Var Significance

Within the scope of **Multiple Linear regression - the t-Test** is conducted for **Individual Var Significance.**

```
Call:
lm(formula = LungCap_cc ~ Age_years + Height_inches)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4080 -0.7097 -0.0078  0.7167  3.1679

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -11.747065   0.476899 -24.632  < 2e-16 ***
Age_years       0.126368   0.017851   7.079 3.45e-12 ***
Height_inches   0.278432   0.009926  28.051  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.056 on 722 degrees of freedom
Multiple R-squared:  0.843,     Adjusted R-squared:  0.8425
F-statistic:  1938 on 2 and 722 DF,  p-value: < 2.2e-16
```

Model Accuracy - **RSE and R-squared**

Residual Standard Error - Due to presence of the Epsilon its calculated as the average amount that the predicted **Response** shall deviate from the **TRUE** regression Line.

---

epsilon (e_t) the Error or Random term -the catch all(measure errors + unknown var's),accounts for all kinds of variability in **Sales**$_t$ which can't be explained by the linear effect of -*p*-indep var's.

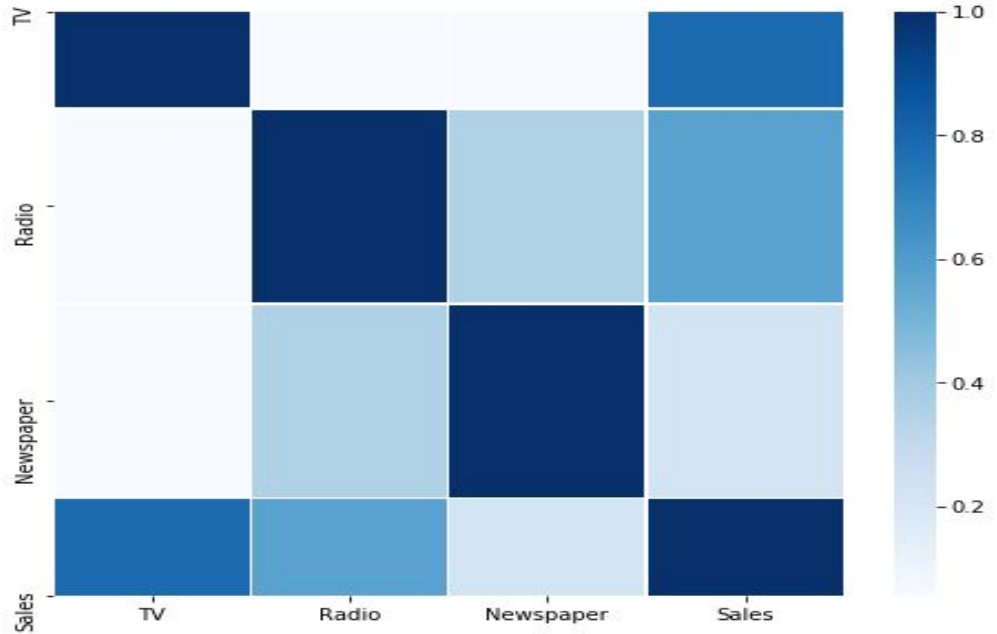
Even when we know the True Regression line and the Populations Params B_0 and B_1 ,we will not be able to predict the True relation between X and Y.
If the RSE value is - 4.5 we can approximate and state that the **Actual Sales vs Predicted sales** differ by - 4.5 Units for the given time period -t.

R-squared and Adjusted R-squared- *as seen in slides above.*

This is Model accuracy not the Best Model for the job - which is comparison between models - AIC , BIC etc ***(not in current scope)***

Multiple Linear Regression - MMM

Py Code

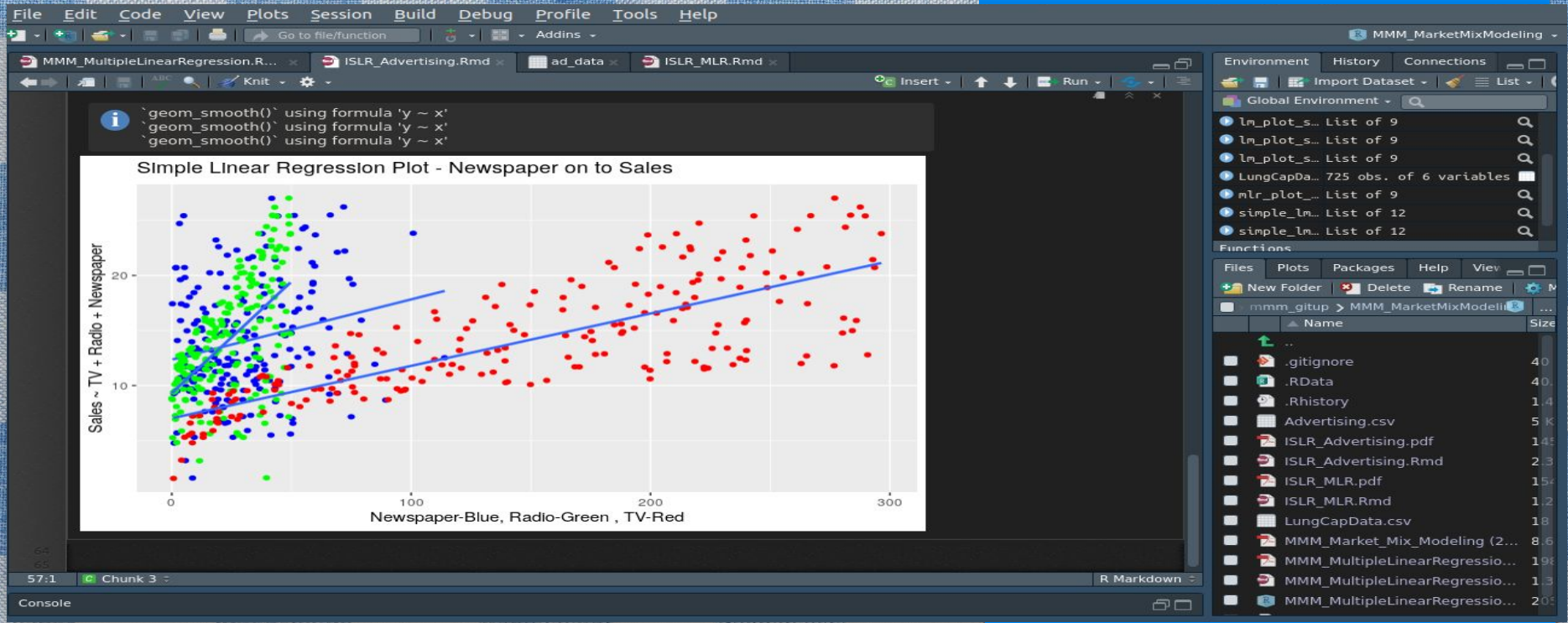


Correlation Plot - *Independent variables* along with *Dependent variable*

(Jupyter Notebook)

Multiple Linear Regression

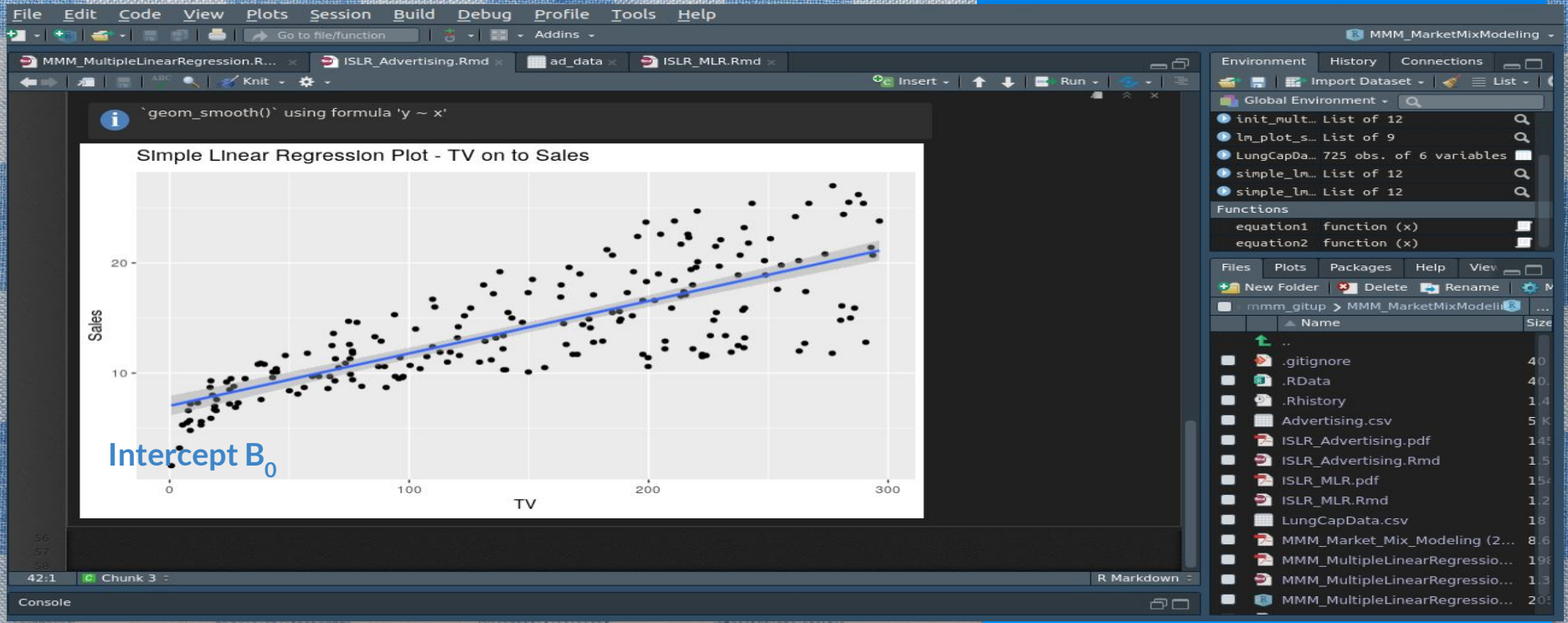
Sales ~ TV + Radio + Newspaper



Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

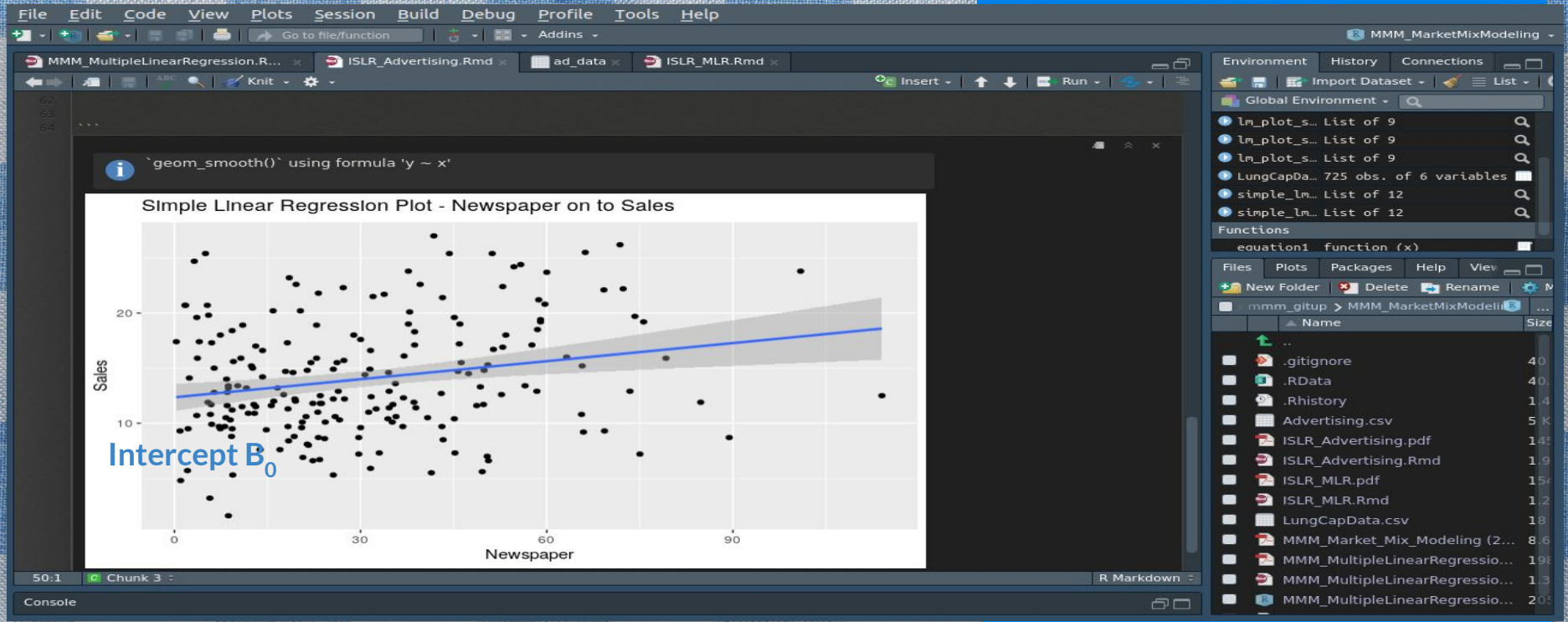
Simple Linear Regression Plot - Tv onto Sales



Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

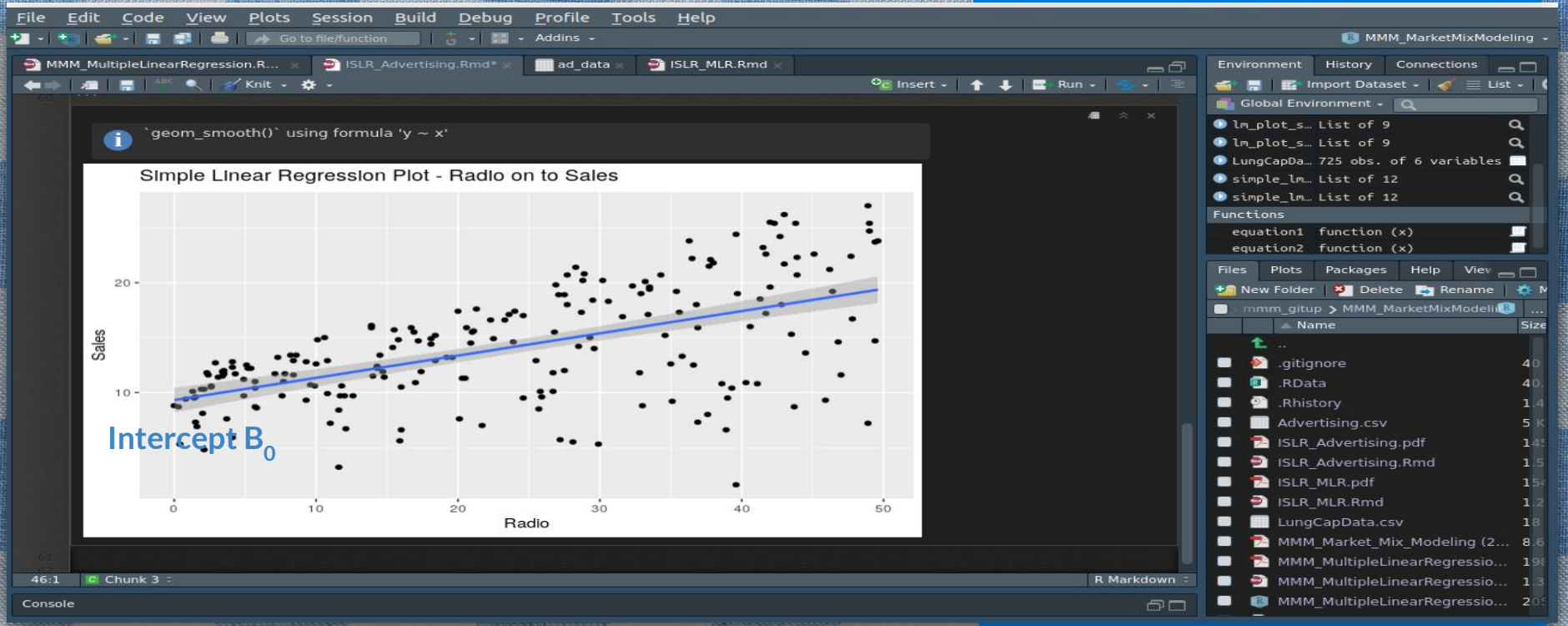
Simple Linear Regression Plot - Newspaper onto Sales



Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

Simple Linear Regression Plot - Radio onto Sales



Digital Cognition(R)

(https://github.com/digital-cognition-co-in)

Multicollinearity with Variance Inflation Factors (VIF)

The issues that Arise from - **Multicollinearity** can be resolved with the Ridge Regression Approach - more on that soon ..
https://github.com/digital-cognition-co-in/MMM_MarketMixModeling/issues/2

Our References :-

OLS - Ordinary Least Squares = https://en.wikipedia.org/wiki/Ordinary_least_squares

from statsmodels.formula.api import ols =
https://scipy-lectures.org/packages/statistics/index.html#multiple-regression-including-multiple-factors

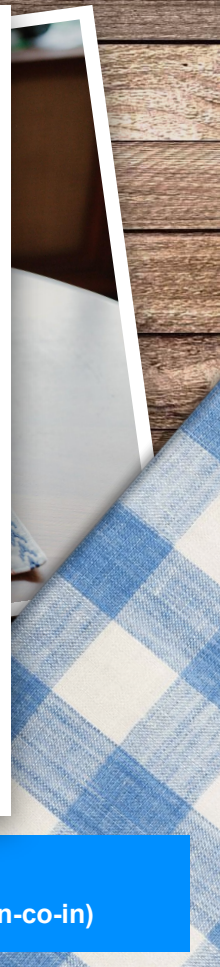
ISLR - An Introduction to Statistical Learning - with Applications in R -
http://faculty.marshall.usc.edu/gareth-james/ISL/

Cornell Univ -
http://mezeylab.cb.bscb.cornell.edu/labmembers/documents/supplement%205%20-%20multiple%20regression.pdf

Media TRP - https://marketing-dictionary.org/t/target-rating-point/

Gross Rating Points or GRPs - https://www.srds.com/frontMatter/sup_serv/calculator/grp_trp/grps_trps.html
https://mpwmarketing.com/2017/05/efficiently-media-buy-tv/

MSE and MSR - https://www.britannica.com/science/mean-square-due-to-error
https://online.stat.psu.edu/stat501/book/export/html/901



Digital Cognition(R)

QnA

Rohan Mathur:

https://www.linkedin.com/in/rohanmathur17/

Rohit Dhankar:

https://www.linkedin.com/in/rohitdhankar/