Graphical user interface, application

Description automatically generatedGraphical user interface, chart, line chart

Description automatically generated

Ggplot resources: facet\_grid

<https://www.datacamp.com/community/tutorials/facets-ggplot-r?utm_source=adwords_ppc&utm_medium=cpc&utm_campaignid=14051819510&utm_adgroupid=&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=x&utm_adpostion=&utm_creative=&utm_targetid=&utm_loc_interest_ms=&utm_loc_physical_ms=9009670&gclid=Cj0KCQiAieWOBhCYARIsANcOw0yHRF7JE7KgttAM4r9FDzbaL-go5T7ZH9NcAAN_gzgx3_mHZ878dboaAutrEALw_wcB>

<http://zevross.com/blog/2019/04/02/easy-multi-panel-plots-in-r-using-facet_wrap-and-facet_grid-from-ggplot2/>

<http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>

Regression resources:

<https://www.statmethods.net/stats/regression.html>

<https://cran.r-project.org/web/packages/olsrr/vignettes/variable_selection.html>

<https://365datascience.com/tutorials/statistics-tutorials/r-squared/>

**12/15/21 Meeting Notes**

Cat1: RSDI

Cat2: Modified differential emotions scales (Emotions)

Cat3: the buss-Perry aggression questionnaire (Aggression)

1. Breakdown the RSDI survey (cat 1) into 3 subcat: physical, functional, emotional
2. Make the same plot again – remove R and L endo
3. Same plot for the subset of people with CT scores
4. See if Cat 2 and Cat 3 helps explain objective scores (CT and endo scores) better than RSDI score

* Learn about the surveys on Wikipedia
* Part 1: correlation between Cat2/Cat3 and RSDI subcategories
* Part 2: regression model to see which variables explain better

1. Surgery: include those before survey & people outside of the 3-month period

* Make a separate plot
* Get the list of people with surgery but no date 🡪 give to Garret

1. Garret’s dataset

# To-Do:

1. Regression diagnostic
2. Different regression model decision: linear, lasso, ridge, elastic-net

# Scatterplots of all variables

Calendar

Description automatically generated

# Removing outliers

Chart

Description automatically generated with low confidence

Remove all num\_allergies outliers (1.5\*IQR below and above 25th and 75th percentile, respectively)

A picture containing diagram

Description automatically generated

Remove only those with num\_allergies >25

Calendar

Description automatically generated

# Scatterplots of obs w/ CT scores

Calendar

Description automatically generated with low confidence

# Fitting Models

## Checking for assumptions

1. Linearity

* **Residual vs Fitted plot:** don’t want any pattern around the red line

1. Independence: predictors are independent

* **Durbin-Watson statistic**: a test for autocorrelation in a regression model's output
* Null hypothesis states that the errors are not auto-correlated with themselves (they are independent)
* p-value > 0.05: evidence to state that our independence assumption is met

1. Residual errors have mean 0

* **Residual vs Fitted plot**: want red line flat on 0

1. Residual errors have constant variance

* **Scale-Location plot**: fitted values vs the square root of the standardized residuals
* Want to see the residual points equally spread around the red line

1. Residual errors are independent from each other

* Understanding of the study design

## Model 1: endo\_score ~ all others

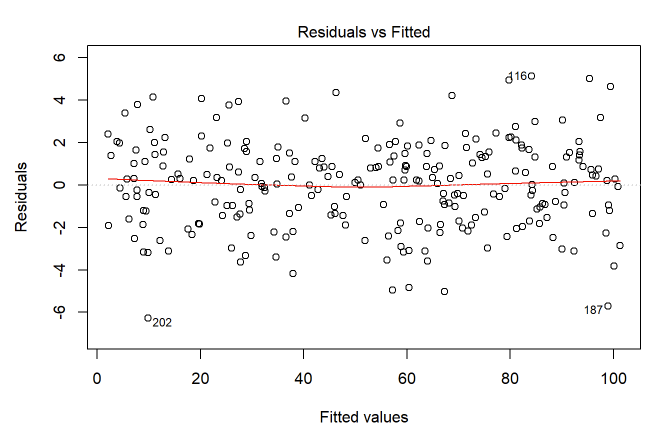
Text

Description automatically generated

Chart, scatter chart

Description automatically generated

[this is what we want to see]:



Chart, scatter chart

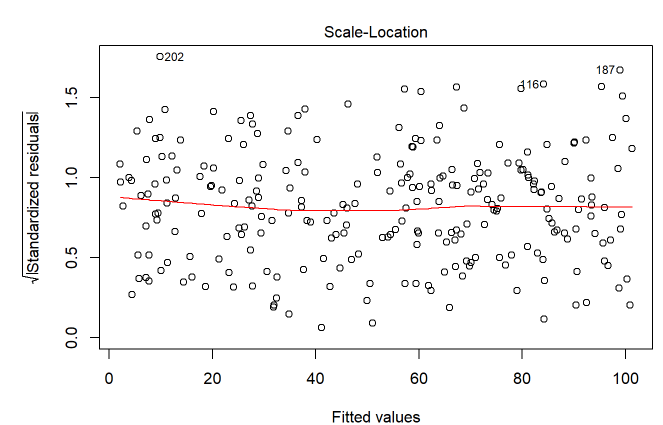
Description automatically generated

Sqrt transformed y value to fix the non-constant variance problem

Chart, scatter chart

Description automatically generated

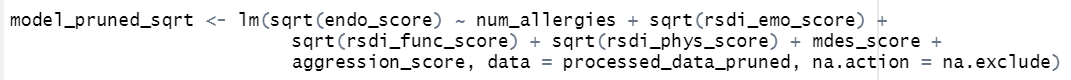
[this is what we want to see]:



Text, letter

Description automatically generatedTable

Description automatically generated

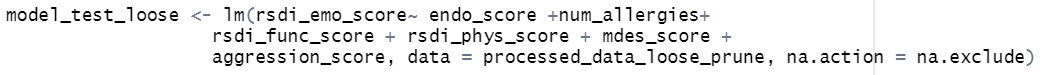


Text

Description automatically generated with medium confidence Table

Description automatically generated

## Model 2: rsdi\_emo\_score ~ all others



Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Sqrt(y) transformation

Chart, scatter chart

Description automatically generated

Table

Description automatically generated with low confidence Table

Description automatically generated

# Histograms: before transformation

## processed\_data\_pruned

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

# Histogram: Transformed data

* Log\_endo\_score
* Log\_rsdi\_emo\_score
* Log\_rsdi\_func\_score
* Log\_rsdi\_phy\_score
* (use the original mdes\_score and aggression\_score)

Chart, histogram

Description automatically generated

Chart, histogram, box and whisker chart

Description automatically generated