fishRprice package

Michael Dowd, Chutimon Singhakarn, Corey Nevels

Summary

The goal of this package:

- To sort and columnize the measurement data (fork length, total length, weight carcass, weight liver)
- To analyzing measured data with sex, age, and season of collection, and visualizing the data in both space and time.

Data description

- The Fish_Data dataset is made of 10,480 observations, measuring fork length, total length, carcass weight, and liver weight of 2,620 individuals collected around Oahu from 2001 to 2021.
- Half of the individuals are Big-eyed scad and half are Bluestriped Snappers.
- This data set follows a non-standard structure, with each individual having 4 lines with the same specimen ID, species ID, Date of Collection, Age, Sex, and Waypoint, but with a different measurement and measurement value, as described previously.

	A	B	C	D	E	F	G	H	I
1	CaseAcc	Animal	DateCollected	Age	Sex	waypoint	Measurement	MeasureValue	MeasurementUnit
2	25174-21	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Fork length	22.5	cm
3	25174-21	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Total length	23.9	cm
4	25174-21	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight carcass	201.1	g
5	25174-21	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight liver	1.489	g
6	25174-27	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Fork length	21.6	cm
7	25174-27	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Total length	22.8	cm
8	25174-27	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Weight carcass	176.7	g
9	25174-27	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Weight liver	0.982	g
10	25174-28	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Fork length	24.7	cm
11	25174-28	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Total length	26	cm
12	25174-28	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight carcass	270	g
13	25174-28	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight liver	1.462	g
14	25174-30	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Fork length	22	cm
15	25174-30	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Total length	23.2	cm
16	25174-30	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight carcass	187.7	g
17	25174-30	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight liver	0.914	g
18	25174-26	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Fork length	20.7	cm
19	25174-26	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Total length	21.7	cm
20	25174-26	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight carcass	148.3	g
21	25174-26	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight liver	0.649	g
22	25174-25	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Fork length	22	cm
23	25174-25	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Total length	23.4	cm
24	25174-25	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight carcass	189.9	g
25	25174-25	Snapper: Bluestriped	16/3/2015	Adult	Male	Barbers Point FR1B	Weight liver	1.432	g
26	25174-24	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Fork length	25.7	cm
27	25174-24	Snapper: Bluestriped	16/3/2015	Adult	Female	Barbers Point FR1B	Total length	27.2	cm

Data wrangling

1. **subset_by_col_value**

This function takes only the observations in a table that match the specified column value in the specified column

Ex. subset_by_col_value(snapper, MeasureValue, ‘Fork length’)

2. **remove_col**

This function removes specified columns

Ex. remove_col(snapper, c(“MeasureValue”)

3. **keep_col**

This function keeps specific columns

Ex. keep_col(snapper, c(“MeasureValue”)

Console | Terminal | Jobs

~/Desktop/MBIO612-Fun R/R project/

```
> library(fishRprice)
> data("scad")
> head(scad)
   CaseAcc         Animal DateCollected   Age  Sex           Location    Measurement MeasureValue MeasurementUnit
1  25142-9 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B    Fork length       25.000               cm
2  25142-9 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B   Total length       27.700               cm
3  25142-9 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B Weight carcass      250.100                g
4  25142-9 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B   Weight liver        3.158                g
5 25142-10 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B    Fork length       25.200               cm
6 25142-10 Scad: Big-eyed       1/15/15 Adult Male Barbers Point FR1B   Total length       27.900               cm
> scad_1 <- remove_col(scad,"MeasurementUnit")
> scad_2 <- subset_by_col_value(scad_1, Measurement, 'Weight carcass')
> scad_3 <- keep_col(scad_2,c("Animal","DateCollected","Age","Sex","Location","Measurement","MeasureValue"))
> head(scad_3)
         Animal DateCollected   Age    Sex           Location    Measurement MeasureValue
1 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        250.1
2 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        246.8
3 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        240.0
4 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        262.4
5 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        254.0
6 Scad: Big-eyed       1/15/15 Adult Female Barbers Point FR1B Weight carcass        226.9
>
```

Data wrangling

4. Snapper_extracted & scad_extracted

A dataset that contains repeated observations of the same individual, extracted and grouped together

Ex. Turns 4,000 rows from the snapper dataset into 1,000

5. dateR

This function converts the date from numeric to dttm format that R can more easily read and places it in a new column

6. Age_dbl, sex_dbl, waypoint_dbl

This function creates a new column that converts factors into numbers. This is done for age, sex, and waypoint in the datasets

Data Wrangling

```
> head(scad_3)
          Animal DateCollected   Age    Sex           Location    Measurement MeasureValue
1 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        250.1
2 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        246.8
3 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        240.0
4 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        262.4
5 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        254.0
6 Scad: Big-eyed       1/15/15 Adult Female Barbers Point FR1B Weight carcass        226.9
> scad_4 <- dateR(scad_3, DateCollected, "%m/%d/%Y")
> head(scad_4)
          Animal DateCollected   Age    Sex           Location    Measurement MeasureValue      DateR
1 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        250.1 0015-01-15
2 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        246.8 0015-01-15
3 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        240.0 0015-01-15
4 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        262.4 0015-01-15
5 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B Weight carcass        254.0 0015-01-15
6 Scad: Big-eyed       1/15/15 Adult Female Barbers Point FR1B Weight carcass        226.9 0015-01-15
>
```

Data Wrangling

In [30]:
```
head(snapper_extracted)
```

CaseAcc	Animal	DateCollected	Age	Sex	Waypoint	Fork_length_cm	Total_length_cm	Weight_carcass_g	Weight_liver_g
25174-21	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.5	23.9	201.1	1.489
25174-27	Snapper: Bluestriped	3/16/15	Adult	Female	Barbers Point FR1B	21.6	22.8	176.7	0.982
25174-28	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	24.7	26.0	270.0	1.462
25174-30	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.0	23.2	187.7	0.914
25174-26	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	20.7	21.7	148.3	0.649
25174-25	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.0	23.4	189.9	1.432

In [32]:
```
snapper_extracted1= age_dbl(snapper_extracted, Age)
snapper_extracted2= sex_dbl(snapper_extracted1, Sex)
snapper_extracted3= waypoint_dbl(snapper_extracted2, Waypoint)
head(snapper_extracted3)
```

CaseAcc	Animal	DateCollected	Age	Sex	Waypoint	Fork_length_cm	Total_length_cm	Weight_carcass_g	Weight_liver_g	Age_dbl	Sex_dbl	Waypoint_dbl
25174-21	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.5	23.9	201.1	1.489	2	3	2
25174-27	Snapper: Bluestriped	3/16/15	Adult	Female	Barbers Point FR1B	21.6	22.8	176.7	0.982	2	2	2
25174-28	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	24.7	26.0	270.0	1.462	2	3	2
25174-30	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.0	23.2	187.7	0.914	2	3	2
25174-26	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	20.7	21.7	148.3	0.649	2	3	2
25174-25	Snapper: Bluestriped	3/16/15	Adult	Male	Barbers Point FR1B	22.0	23.4	189.9	1.432	2	3	2

Data Wrangling

```
> head(scad_full)
  X  CaseAcc         Animal DateCollected   Age    Sex           Waypoint Fork_length_cm Total_length_cm Weight_carcass_g Weight_liver_g
1 1  25142-9 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.0            27.7            250.1          3.158
2 2 25142-10 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.2            27.9            246.8          2.201
3 3  25142-1 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.0            27.1            240.0          2.233
4 4  25142-8 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.9            28.4            262.4          2.684
5 5  25142-7 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.6            28.2            254.0          2.758
6 6  25142-6 Scad: Big-eyed       1/15/15 Adult Female Barbers Point FR1B           24.5            26.9            226.9          2.801
  Waypoint_dbl Age_dbl Sex_dbl      DateR
1            1       1       2 0015-01-15
2            1       1       2 0015-01-15
3            1       1       2 0015-01-15
4            1       1       2 0015-01-15
5            1       1       2 0015-01-15
6            1       1       1 0015-01-15
> scad_6 <- keep_col(scad_full,c("Animal","DateR","Age_dbl","Sex_dbl","Waypoint_dbl",
+                           "Fork_length_cm","Total_length_cm","Weight_carcass_g","Weight_liver_g"))
> head(scad_6)
         Animal      DateR Age_dbl Sex_dbl Waypoint_dbl Fork_length_cm Total_length_cm Weight_carcass_g Weight_liver_g
1 Scad: Big-eyed 0015-01-15       1       2            1           25.0            27.7            250.1          3.158
2 Scad: Big-eyed 0015-01-15       1       2            1           25.2            27.9            246.8          2.201
3 Scad: Big-eyed 0015-01-15       1       2            1           25.0            27.1            240.0          2.233
4 Scad: Big-eyed 0015-01-15       1       2            1           25.9            28.4            262.4          2.684
5 Scad: Big-eyed 0015-01-15       1       2            1           25.6            28.2            254.0          2.758
6 Scad: Big-eyed 0015-01-15       1       1            1           24.5            26.9            226.9          2.801
>
```

Exploratory Data Analysis

1. **get_stats**

This function gives specified statistic based on the 2 parameters supplied

get_stats(scad_extracted, ‘Weight_liver_g’, ‘Age’, sd)

Gives the standard deviation of column ‘Weight_liver_g’ per ‘Age’ category

2. **HSI**

This function creates a new column with the Hepatosomatic index (the ratio of liver weight to total body weight)

3. **Cond_factor_k**

This function creates a new column that gives the Fulton condition factor, K, for each observation

$K = 100 \times weight/length\text{^}3$ - the BMI or health index of the individual fish

Exploratory Data Analysis

Console | Terminal | Jobs

~/Desktop/MBIO612-Fun R/R project/

```
> get_stats(scad_full,'Weight_carcass_g','Age',mean)
   Adult  Unknown
249.7199 173.7667
> get_stats(scad_full,'Weight_carcass_g','Age',sd)
  Adult Unknown
66.7333 69.8702
> get_stats(scad_full,'Total_length_cm','Age',mean)
   Adult  Unknown
27.50930 24.90333
> get_stats(scad_full,'Total_length_cm','Age',sd)
   Adult  Unknown
2.259162 3.058001
> scad_full_HSI <- HSI(scad_full,Weight_liver_g,Weight_carcass_g)
> scad_full_HSI_K <- cond_factor_k(scad_full_HSI,Weight_carcass_g,Total_length_cm)
> head(scad_full_HSI_K)
  X  CaseAcc         Animal DateCollected   Age    Sex           Waypoint Fork_length_cm Total_length_cm Weight_carcass_g Weight_liver_g Waypoint_dbl Age_dbl
1 1  25142-9 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.0            27.7            250.1          3.158            1       1
2 2 25142-10 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.2            27.9            246.8          2.201            1       1
3 3  25142-1 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.0            27.1            240.0          2.233            1       1
4 4  25142-8 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.9            28.4            262.4          2.684            1       1
5 5  25142-7 Scad: Big-eyed       1/15/15 Adult   Male Barbers Point FR1B           25.6            28.2            254.0          2.758            1       1
6 6  25142-6 Scad: Big-eyed       1/15/15 Adult Female Barbers Point FR1B           24.5            26.9            226.9          2.801            1       1
  Sex_dbl      DateR       HSI Fulton_condition_factor_K
1       2 0015-01-15 1.2626949                 1.176723
2       2 0015-01-15 0.8918152                 1.136403
3       2 0015-01-15 0.9304167                 1.205878
4       2 0015-01-15 1.0228659                 1.145536
5       2 0015-01-15 1.0858268                 1.132626
6       1 0015-01-15 1.2344645                 1.165675
>
```

Exploratory Data Analysis

4. show_boxplot

Shows a boxplot for a given column range

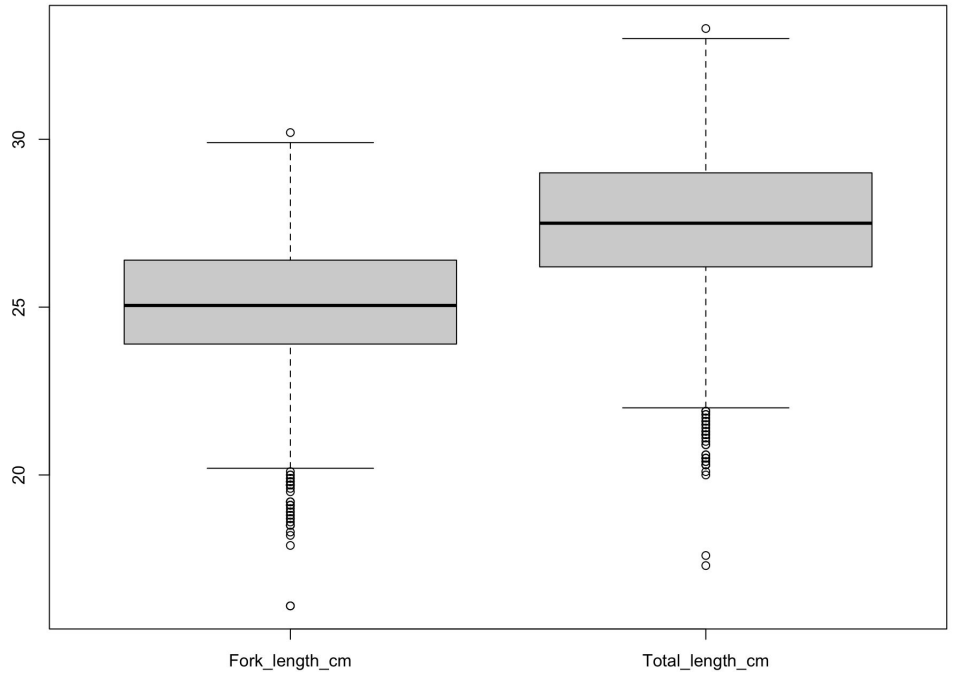
show_boxplot(snapper,) : will show all columns

show_boxplot(scad, 7) : will only show the specified column

show_boxplot(scad, 7:9) will show boxplots for a specified column range

Exploratory Data Analysis

show_boxplot(scad_full_HSI_K,8:9)



Analysis/Modeling

1. linear_model_summary

This function provides a summary of a linear model based on the specified variables

Ex. linear_model_summary(snapper_full, ‘Total_length_cm’, ‘Sex_dbl’)

Analysis/Modeling

Pr (> | t |) describes the significance

Console | Terminal | Jobs

~/Desktop/MBIO612-Fun R/R project/

```
> linear_model_summary(scad_full_HSI_K, 'Total_length_cm', 'Sex_dbl')

Call:
stats::lm(formula = response ~ explanatory)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5414 -1.2097  0.1744  1.5903  6.0744

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.9939     0.1825 153.416  < 2e-16 ***
explanatory  -0.3841     0.1207  -3.183  0.00149 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.304 on 1308 degrees of freedom
Multiple R-squared:  0.007684,  Adjusted R-squared:  0.006926
F-statistic: 10.13 on 1 and 1308 DF,  p-value: 0.001494

>
```

Analysis/Modeling

2. two_explain_mlm_summary

Provides a summary of a linear model based on 2 specified variables

two_explan_mlm_significance(snapper_full, ‘Total_length_cm’, ‘Sex_dbl’, ‘Age_dbl’)

3. three_explan_mlm_summary

Provides a summary of a liner model based on 3 specified variables

three_explan_mlm_summary(scad_full, ‘Total_length_cm’, ‘Sex_dbl’, ‘Age_dbl’, ‘Waypoint_dbl’)

Exploratory Data Analysis

Returns a summary of the linear model

Pr (> | t |) describes the significance

Someone who doesn't know R can use this and easily get a linear model summary for their data

Console | Terminal | Jobs

~/Desktop/MBIO612-Fun R/R project/

```
> two_explan_mlm_summary(scad_full_HSI_K, 'Total_length_cm', 'Sex_dbl', 'Age_dbl')

Call:
stats::lm(formula = response ~ explanatory1 + explanatory2)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6391 -1.2582  0.1014  1.6014  6.0014

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   30.5808     0.4606  66.396  < 2e-16 ***
explanatory1  -0.3595     0.1191  -3.018  0.00259 **
explanatory2  -2.5631     0.4200  -6.102 1.38e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.273 on 1307 degrees of freedom
Multiple R-squared:  0.03517,   Adjusted R-squared:  0.03369
F-statistic: 23.82 on 2 and 1307 DF,  p-value: 6.894e-11

> three_explan_mlm_summary(scad_full_HSI_K, 'Total_length_cm', 'Sex_dbl', 'Age_dbl', 'Waypoint_dbl')

Call:
stats::lm(formula = response ~ explanatory1 + explanatory2 +
    explanatory3)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6471 -1.2743  0.0675  1.4353  6.2545

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  29.41942    0.47619  61.781  < 2e-16 ***
explanatory1 -0.34747    0.11664  -2.979  0.00295 **
explanatory2 -2.31689    0.41251  -5.617 2.38e-08 ***
explanatory3  0.17740    0.02338   7.588 6.13e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.225 on 1306 degrees of freedom
Multiple R-squared:  0.07592,   Adjusted R-squared:  0.07379
F-statistic: 35.76 on 3 and 1306 DF,  p-value: < 2.2e-16
```

Analysis/Modeling

4. two_explain_mlm_significance

Provides the cosignificance of a linear model based on 2 specified variables

two_explan_mlm_significance(scad_full, ‘Total_length_cm’, ‘Sex_dbl’, ‘Age_dbl’)

5. three_explan_mlm_significance

Provides the cosignificance of a linear model based on 3 specified variables

three_explan_mlm_significance(snapper_full, ‘Total_length_cm’, ‘Sex_dbl’, ‘Age_dbl’, ‘Waypoint_dbl’)

Exploratory Data Analysis

Returns a stepAIC of the explanatory variables.

The variables displayed with the smallest AIC are significant

Console | Terminal | Jobs

~/Desktop/MBIO612-Fun R/R project/

```
> two_explan_mlm_significance(scad_full_HSI_K, 'Total_length_cm', 'Sex_dbl', 'Age_dbl')
Start:  AIC=2154.14
response ~ explanatory1 + explanatory2

               Df Sum of Sq    RSS    AIC
<none>                      6751.9 2154.1
- explanatory1  1    47.058 6799.0 2161.2
- explanatory2  1   192.348 6944.3 2188.9

Call:
stats::lm(formula = response ~ explanatory1 + explanatory2)

Coefficients:
 (Intercept)  explanatory1  explanatory2
     30.5808       -0.3595       -2.5631

> three_explan_mlm_significance(scad_full_HSI_K, 'Total_length_cm', 'Sex_dbl', 'Age_dbl', 'Waypoint_dbl')
Start:  AIC=2099.61
response ~ explanatory1 + explanatory2 + explanatory3

               Df Sum of Sq    RSS    AIC
<none>                      6466.8 2099.6
- explanatory1  1     43.94 6510.7 2106.5
- explanatory2  1    156.20 6623.0 2128.9
- explanatory3  1    285.14 6751.9 2154.1

Call:
stats::lm(formula = response ~ explanatory1 + explanatory2 +
    explanatory3)

Coefficients:
 (Intercept)  explanatory1  explanatory2  explanatory3
     29.4194       -0.3475       -2.3169        0.1774

>
```

Thank you