ggplotBasics

Dieter

2022-08-26

Read in some data

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data <-read_csv('data/body.csv')
```

```
## Rows: 507 Columns: 25
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (25): Biacromial, Biiliac, Bitrochanteric, ChestDepth, ChestDia, ElbowDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Revisit the simple linear model

```
model <- lm(data$Weight ~ data$Waist)
summary(model)
```

```
##
## Call:
## lm(formula = data$Weight ~ data$Waist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.3211  -3.5995  -0.0936   3.6932  19.0536
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.18382    1.79291  -8.469 2.72e-16 ***
## data$Waist    1.09550    0.02306  47.514  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.712 on 505 degrees of freedom
## Multiple R-squared:  0.8172, Adjusted R-squared:  0.8168
## F-statistic:  2258 on 1 and 505 DF,  p-value: < 2.2e-16
```

```
model <- lm(data$Weight ~ data$Waist * data$Shoulder)
summary(model)
```

```
##
## Call:
## lm(formula = data$Weight ~ data$Waist * data$Shoulder)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.597  -2.938  -0.363   3.077  15.088
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -15.998345  16.312526  -0.981   0.3272
## data$Waist                 0.336581   0.221661   1.518   0.1295
## data$Shoulder              0.311073   0.149149   2.086   0.0375 *
## data$Waist:data$Shoulder   0.003037   0.001944   1.562   0.1189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.752 on 503 degrees of freedom
## Multiple R-squared:  0.874,  Adjusted R-squared:  0.8732
## F-statistic:  1163 on 3 and 503 DF,  p-value: < 2.2e-16
```

```
model1 <- lm(data$Weight ~ data$Waist)
model2 <- lm(data$Weight ~ data$Waist + data$Shoulder)

summary(model2)
```

```
##
## Call:
## lm(formula = data$Weight ~ data$Waist + data$Shoulder)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.8730  -3.1827  -0.2748   3.0894  15.7445
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -41.22589    2.29507  -17.96   <2e-16 ***
## data$Waist      0.67877    0.03386   20.05   <2e-16 ***
## data$Shoulder   0.53720    0.03594   14.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.759 on 504 degrees of freedom
## Multiple R-squared:  0.8733, Adjusted R-squared:  0.8728
## F-statistic:  1738 on 2 and 504 DF,  p-value: < 2.2e-16
```

```
anova(model1, model2)
```

```
## Analysis of Variance Table
```

```
##
## Model 1: data$Weight ~ data$Waist
## Model 2: data$Weight ~ data$Waist + data$Shoulder
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    505 16475
## 2    504 11415  1    5059.8 223.41 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```