Vis-MLM data sets, analyses & plots

This document is simply notes-to-self on data sets I plan to use in the book, and others I might use, with links to examples I will draw on. **TODO**: Flesh out, and assign them to chapters.

## Penguins

The iris data is probably too tired. The palmerpenguins data set is an easy replacement. The book will use a slightly cleaned-up version, peng. Here are a few use cases for examples.

* Penguin data: Multivariate EDA: <https://rpubs.com/friendly/penguin-biplots>
* Penguins: MANOVA and HE plots: <https://rpubs.com/friendly/penguin-manova>
* See also: <https://www.r-bloggers.com/2020/07/basic-data-analysis-with-palmerpenguins/>
* Penguins go parallel: <https://www.tandfonline.com/doi/epdf/10.1080/10618600.2023.2195462>
* <https://cameronpatrick.com/post/2023/06/dplyr-fitting-multiple-models-at-once/>
* Uncertainty: <https://github.com/mjskay/uncertainty-examples/blob/master/penguins.qmd>

## Guerry

Andre-Michel Guerry’s data on *Moral Statistics of France* provides several data sets of rich historical interest. The main data set contains variables reflecting personal crime, property crime, literacy, suicides, children born out of wedlock, … These are available in the [Guerry](https://github.com/friendly/Guerry) package. The package vignette [guerry-multivariate](https://rdrr.io/cran/Guerry/f/vignettes/guerry-multivariate.Rmd) provides detailed examples of the topics mentioned below.

* Bivariate: personal & property crime
* Multivariate
  + parallel coords plot –
  + radar plot of means
  + Biplot
  + MANOVA / candisc

## Lahman

Baseball data from the [Lahman](https://CRAN.R-project.org/package=Lahman) package gives a rich set of variables to explore a variety of multivariate questions and analyses.

## heplots data sets

The [heplots](https://CRAN.R-project.org/package=heplots) package provides many data sets I plan to use, all with detailed examples and vignettes in the package.

## carData

The [carData](https://CRAN.R-project.org/package=heplots) package provides a nice collection of simple data sets for illustrating graphical methods and regression diagnostics.

### Duncan data

* See 6140, [Regression diagnostics](http://euclid.psych.yorku.ca/www/psy6140/lectures/RegDiagnostics2x2.pdf)

### prestige data

* Examples from Vis-MLM-Course, Lecture 1 C:-course{prestige.R, prestige-ggplot.R}
* Regression Model for the Prestige Level of Occupations, <http://rstudio-pubs-static.s3.amazonaws.com/425420_448c3a57871f4ac3a98f7b7781ffc91e.html>

### Salaries

+ https://rkabacoff.github.io/datavis/Multivariate.html

## Others

* <https://www.r-bloggers.com/2021/11/manovamultivariate-analysis-of-variance-using-r/>

## Spread-level plots & transformations to symmetry

* <https://mgimond.github.io/ES218/sl_plot.html>
* <https://mgimond.github.io/ES218/re_express.html>
* <https://book.stat420.org/transformations.html>