



Visualize This!

Using R in Introductory Statistics Courses

Teaching Statistics: Start With the Data



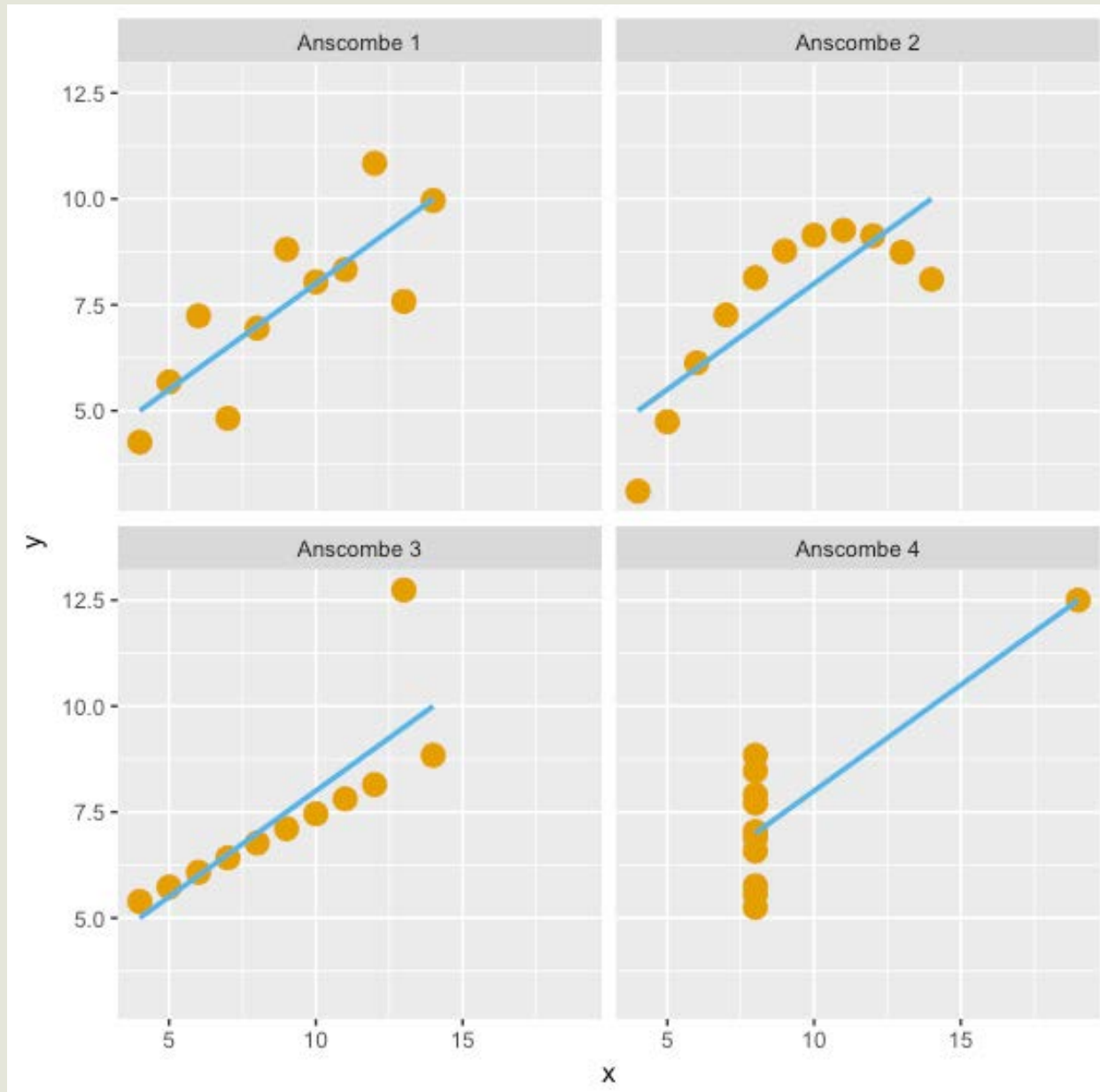
Image and other content provided by “the Open Knowledge Foundation” by creative commons license
<https://blog.okfn.org/2013/10/03/defining-open-data/>



No Installation Required – Why R ?

- Fast becoming the most utilized tool in statistical analysis
- Costs nothing to use
- **Considered to be the best visualization tool**
- Growing user support and development community
- Data literacy
- Lots of Code Examples require no programming expertise

Anscombe's Data Plots in R





Anscombe's Quartet – Results of Linear Regression

Property	Value	R code
Mean of x	9	<code>mean(anscombe\$x1)</code>
Sample variance of x	11	<code>var(anscombe\$x1)</code>
Mean of y	7.50	<code>mean(anscombe\$y1)</code>
Sample variance of y	4.125	<code>var(anscombe\$y1)</code>
Correlation between x and y	0.816	<code>cor(anscombe\$x1, anscombe\$y1)</code>
Linear regression line	$y = 3.00 + 0.500x$	R Code Block
Coefficient of determination of the linear regression	0.67	R Code Block

Repeat these for x_1, x_2, x_3, x_4 and y_1, y_2, y_3, y_4

The background features a dark grey, chalkboard-like texture with various white line drawings of scientific and mathematical symbols. These include a globe, a microscope, a book, a percentage sign, and various geometric shapes and arrows.

DatasauRus

More Fun With Nonlinear Data



DatasauRus Dozen Activity

R Code

- jupyter notebooks
- Built-in

Copy-Paste

- Snippets
- RStudio
- Click “Run”

Save Image

- Image plots
- Drag to desktop
- png file

Discuss

- More time for discussion in class



away

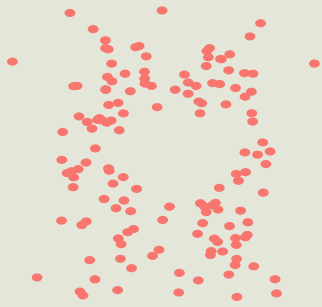
bullseye

circle

dino

dots

away



h lines

high lines

slant down

slant up

star_{star}

v lines

wide lines

x shape

Each of these has the same regression parameters

R Code Blocks – Anscombe Data

```
# The following will produce the 4 plots which
# can be exported in a single image file
x <- c(anscombe$x1,anscombe$x2,anscombe$x3,anscombe$x4)
y <- c(anscombe$y1,anscombe$y2,anscombe$y3,anscombe$y4)
dataset <- c(1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,
            3,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,4)
anscombe1 <- as.data.frame(cbind(x,y,dataset))
names(anscombe1) <- c("x", "y", "dataset")
anscombe1$dataset <- factor(anscombe1$dataset,levels = 1:4,
                          labels = sprintf("Anscombe %s",1:4))
ggplot(anscombe1, aes(x=x, y=y)) +
  geom_point(color='#E69F00', size = 4) +
  geom_smooth(method=lm, se=FALSE, color = '#56B4E9') +
  facet_wrap(~dataset, ncol=2)
```

R Code Blocks – Anscombe Data

```
# Using the anscombe data, find the regression lines
# and related summary statistics
require(stats); require(graphics)
summary_data <- summary(anscombe)
ff <- y ~ x
mods <- setNames(as.list(1:4), paste0("lm", 1:4))
for(i in 1:4) {
  ff[2:3] <- lapply(paste0(c("y","x"), i), as.name)
  mods[[i]] <- lmi <- lm(ff, data = anscombe)
  print(anova(lmi))
}
regression_coeff <- sapply(mods, coef)
print(regression_coeff)
lapply(mods, function(fm) coef(summary(fm)))
```

References – Where To Find Materials

- Anscombe's Quartet: Anscombe, F.J. (1973). "Graphs in Statistical Analysis". *The American Statistician*. The American Statistician, Vol. 27, No. 1. **27** (1) pages 17-21. JSTOR 2682899. doi:10.2307/2682899
- To create the visualizations using the "Datasaurus Dozen" the R Code is in a jupyter notebook and is publicly available. Please take care NOT to alter the code:
https://rnotebook.io/anon/77500e35422c9834/notebooks/datasauRus_dozen_plots.ipynb
- Snippets is an online, web-based tool that allows one to run R code and create visualizations in the browser. Graphs and plots can be dragged to the desktop and will save as an image file (.png) <https://rdr.io/snippets/>
- R Datasets package: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html>
- R programming language is free and installed via "precompiled binary distributions"; click on "Download R for (Mac) OS X" or "Download R for Windows" as appropriate from this site:
<http://cran.utstat.utoronto.ca/>
- Rstudio is a more user-friendly way of working with R. You need to download R first, then Rstudio Desktop:
<https://www.rstudio.com/products/rstudio/download/>
- Want to learn as an absolute beginner? Try this free online introduction:
<https://www.datacamp.com/courses/free-introduction-to-r>
- DataCamp Classroom is FREE:
<https://www.datacamp.com/groups/education>