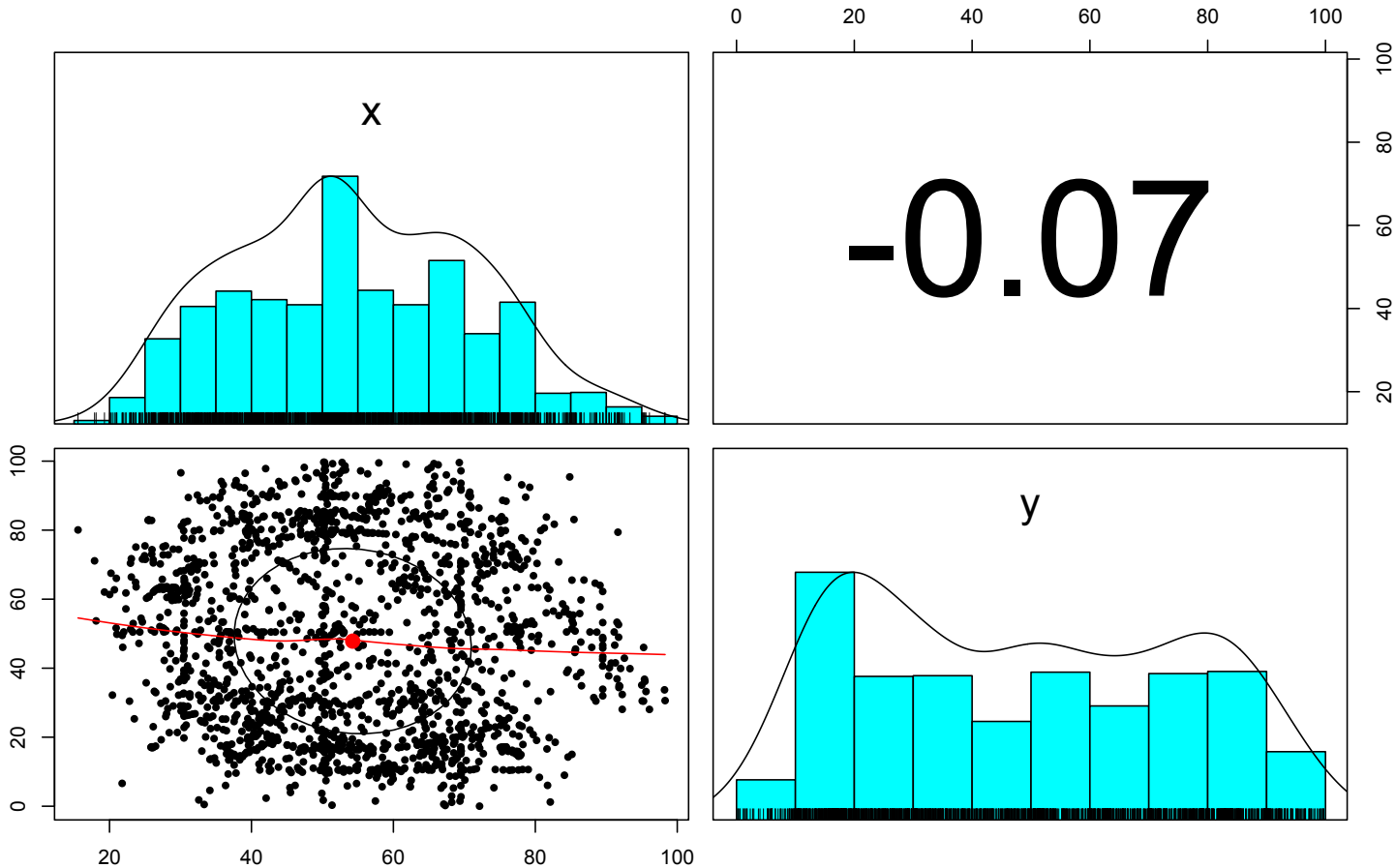
# Psychology 350: Special Topics
# An introduction to R for psychological research

William Revelle
Northwestern University
Evanston, Illinois USA

https://personality-project.org/courses/350


Spring, 2024

# A rather boring data set.
# The overall plot of all the data shows no relationship




| | vars | n | mean | sd | median | trimmed | mad | min | max | range | skew | kurtosis | se |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dataset* | 1 | 1846 | 7.00 | 3.74 | 7.00 | 7.00 | 4.45 | 1.00 | 13.00 | 12.00 | 0.00 | -1.22 | 0.09 |
| x | 2 | 1846 | 54.27 | 16.71 | 52.59 | 54.03 | 19.27 | 15.56 | 98.29 | 82.73 | 0.14 | -0.70 | 0.39 |
| y | 3 | 1846 | 47.84 | 26.85 | 47.59 | 47.18 | 36.73 | 0.02 | 99.69 | 99.68 | 0.16 | -1.28 | 0.62 |

# Correlations need to be seen to understand the data

From Davies R, Locke S, D'Agostino McGowan L (2022). datasauRus: Datasets from the Datasaurus Dozen. R package version 0.1.6, <https://CRAN.R-project.org/package=datasauRus>.

R code

```
library(dinosaRus)  #install it first and then make it active
#use  functions from the psych package:
describe(datasaurus_dozen)   #basic descriptive statisitics
pairs.panels(datasaurus_dozen[2:3])  #a splom of the data

#now break it down by group
sb <- statsBy(x = datasaurus_dozen[2:3],
         group =datasaurus_dozen$dataset,
         cors=TRUE)
#grab the results and organize them into a data.frame
df <- data.frame(mean=sb$mean,sd=sb$sd,cor=sb$within)
rownames(df) <- datasaurus_dozen$dataset)
round(df,2) #round them to make it pretty
```

```
           mean.x mean.y mean.group  sd.x  sd.y sd.group   x.y
away         54.27  47.83          1 16.77 26.94        0 -0.06
bullseye     54.27  47.83          2 16.77 26.94        0 -0.07
circle       54.27  47.84          3 16.76 26.93        0 -0.07
dino         54.26  47.83          4 16.77 26.94        0 -0.06
dots         54.26  47.84          5 16.77 26.93        0 -0.06
h_lines      54.26  47.83          6 16.77 26.94        0 -0.06
high_lines   54.27  47.84          7 16.77 26.94        0 -0.07
slant_down   54.27  47.84          8 16.77 26.94        0 -0.07
slant_up     54.27  47.83          9 16.77 26.94        0 -0.07
star         54.27  47.84         10 16.77 26.93        0 -0.06
v_lines      54.27  47.84         11 16.77 26.94        0 -0.07
wide_lines   54.27  47.83         12 16.77 26.94        0 -0.07
```

# Plotting the data by group shows something interesting

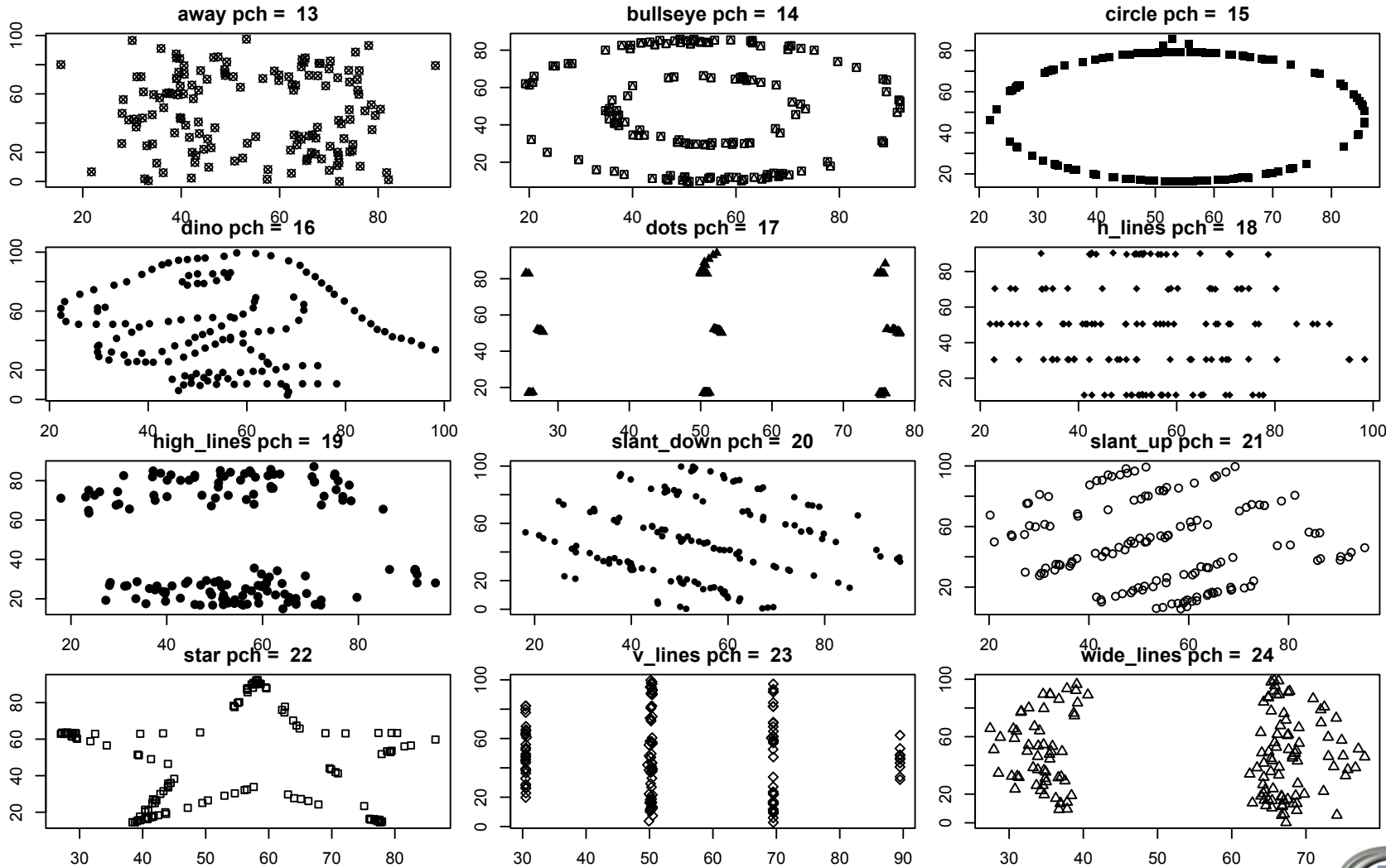
Adapted from the help page for dinosaurus dozen

R code

```
# Base R plots    #use regular plotting function
state = par("mar", "mfrow")   #set the rows, columns and margins
par(mfrow = c(4, 3), mar = c(1, 2, 2, 1))   # 4 rows, 3 columns

 #get the names of the data sets
 sets = sort(unique(datasaurus_dozen$dataset))

#do a loop across the sets
i <- 1   #set a counter so we can change plot character
for (s in sets[1:12]) {           #just show the first 12
 df = datasaurus_dozen[datasaurus_dozen$dataset == s, ]
  plot(df$x, df$y, pch = (12+i))
   title(paste(s, "pch = ", i + 12))
   i <- i + 1    #increment the counter
 } #end of loop

par(state)  #reset settings to the defaults
```

From Davies R, Locke S, D'Agostino McGowan L (2022). datasauRus: Datasets from the Datasaurus Dozen. R package version 0.1.6, <https://CRAN.R-project.org/package=datasauRus>.

# Plotting data is very helpful




All sets $\mu_x = 54.26, \mu_y = 47.84, \sigma_x = 16.77 \sigma_y = 26.93, r = .07$