RAPORT

Mróz & Podsiad

Introduction

Naszym artykułem było "Visualizing Complex Data With Embedded Plots". Artukuł ten miał plik .zip "supplementary" w którym to zostały spakowane międzi innymi 3 pliki .R które miały reprodukować przykłady z artykułu oraz plik .pdf "Appendix" w którym zostały przedstawione rozszerzone użycia pakietu. Mieliśmy jednak z reprodukcją tych przykładów pewien problem, mianowicie:

```
install.packages("ggsubplot")

Installing package into &C:/Users/mrozz/Documents/R/win-library/3.6&

(as &lib is unspecified)

warning in install.packages:
 package 'ggsubplot' is not available (for R version 3.6.3)
```

The Struggle

Próbowaliśmy więc poszukać tego pakietu na CRANie, i natkneliśmy się na

Package 'ggsubplot' was removed from the CRAN repository.

Formerly available versions can be obtained from the archive.

Archived on 2016-01-11 as requested by the maintainer <garrett@rstudio.com>.

Please use the canonical form https://CRAN.R-project.org/package=ggsubplot to link to this page.

Jak widać pakiet ten nie jest już wspierany. Nawet przy próbie zainstalowania poprzednych wersji są duże problemy z kompatybilnością innych pakietów. Dodatkowo natknęliśmy się też na kolejny problem: nie wszystkie datasety się ładowały

```
> # nasa data set loaded with ggsubplot package
> # changing temperature units to fahrenheit
> nasa$fahrenheit <- nasa$surftemp * 9 / 5 - 459.67
Error: object 'nasa' not found
> nasa$atmos.fahrenheit <- nasa$temperature * 9 / 5 - 459.67
Error: object 'nasa' not found
> # subset of data centered on Mexico for Figure 4
> subnasa <- subset(nasa, long < -80 & lat > 10 & long > -107)
Error in subset(nasa, long < -80 & lat > 10 & long > -107):
    object 'nasa' not found
> |
```

Dlatego też jedyne figury które udało nam się zreprodukować to były te figury, które były robione po bożemu w ggplocie i z załadowanymi danymi.

```
Poniżej chunk z załadowanymi danymi które da się załadować.
# O-clean.r
# load and clean data that appears in the figures
library(reshape2)
library(plyr)
library(maps)
##
## Attaching package: 'maps'
## The following object is masked from 'package:plyr':
##
##
      ozone
library(ggplot2)
# install.packages("ggsubplot")
# library(qqsubplot)
###
                      Map backgrounds
                                                         ###
# qetbox by Heike Hoffman, trims map polygons for figure backgrounds
# https://qithub.com/qqobi/paper-climate/blob/master/code/maps.r
getbox <- function (map, xlim, ylim) {</pre>
  # identify all regions involved
 small <- subset(map, (long > xlim[1]) & (long < xlim[2]) & (lat > ylim[1]) & (lat < ylim[2]))</pre>
 regions <- unique(small$region)</pre>
 small <- subset(map, region %in% regions)</pre>
 # now shrink all nodes back to the bounding box
 small$long <- pmax(small$long, xlim[1])</pre>
 small$long <- pmin(small$long, xlim[2])</pre>
 small$lat <- pmax(small$lat, ylim[1])</pre>
 small$lat <- pmin(small$lat, ylim[2])</pre>
 # Remove slivvers
 small <- ddply(small, "group", function(df) {</pre>
   if (diff(range(df$long)) < 1e-6) return(NULL)
```

```
if (diff(range(df$lat)) < 1e-6) return(NULL)</pre>
 })
 small
## Central America for Figure 1-b
## adapted from map_nasa
# https://github.com/ggobi/paper-climate/blob/master/code/maps.r
# assembling data
world <- map_data("world")</pre>
states <- map_data("state")</pre>
states$group <- max(world$group) + states$group</pre>
both <- rbind(world, states)</pre>
americas <- getbox(both, xlim = c(-115, -55), ylim = c(-21.1, 36.6))
# building ggplot2 layer
map_americas <- list(</pre>
  geom_polygon(aes(long, lat, group = group), data = americas, fill = "grey70",
    colour = "grey60", inherit.aes = FALSE, show_guide = FALSE),
  scale_x_continuous("", breaks = NULL, expand = c(0.02, 0)),
  scale_y_continuous("", breaks = NULL, expand = c(0.02, 0)))
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
## Afghanistan for Figures 2 and 3
afghanistan \leftarrow getbox(world, c(60,75), c(28, 39))
map_afghan <- list(</pre>
  geom_polygon(aes(long, lat, group = group), data = afghanistan,
    fill = "grey80", colour = "white", inherit.aes = FALSE,
    show_guide = FALSE),
  scale_x_continuous("", breaks = NULL, expand = c(0.02, 0)),
  scale_y_continuous("", breaks = NULL, expand = c(0.02, 0)))
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
## Mexico and lower US for Figure 4
north\_america \leftarrow getbox(both, xlim = c(-107.5, -80), ylim = c(11, 37.5))
map north <- list(</pre>
  geom_polygon(aes(long, lat, group = group), data = north_america, fill = "grey80",
    colour = "grey70", inherit.aes = FALSE, show_guide = FALSE),
  scale_x_continuous("", breaks = NULL, expand = c(0.02, 0)),
  scale_y_continuous("", breaks = NULL, expand = c(0.02, 0)))
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
## Just the outlines of Mexico and lower US for Figure 4.c
lines_north <- list(</pre>
  geom_polygon(aes(long, lat, group = group), data = north_america, fill = NA,
    colour = "black", inherit.aes = FALSE, show_guide = FALSE),
  scale_x_continuous("", breaks = NULL, expand = c(0.02, 0)),
  scale_y_continuous("", breaks = NULL, expand = c(0.02, 0)))
```

Warning: `show_guide` has been deprecated. Please use `show.legend` instead.

```
## United States for Figure 8
us <- subset(getbox(both, xlim = c(-126, -65), ylim = c(24, 52)),
           region != "Great Lakes")
# making layer
map_us <- list(</pre>
 geom_polygon(aes(long, lat, group = group), data = us, fill = "grey80",
   colour = "grey90", inherit.aes = FALSE, show_guide = FALSE),
 scale_x_continuous("", breaks = NULL, expand = c(0.02, 0)),
 scale_y_continuous("", breaks = NULL, expand = c(0.02, 0)))
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
Mauna Lau Carbon Dioxide Seasonal Trend Data
# retrieve data
# must be connected to internet
# carbon.a <- read.table(header = TRUE, nrow = 32,
# file = "http://www.stat.purdue.edu/~wsc/visualizing.datatables/carbon.dioxide")
# carbon.b <- read.table(header = TRUE, skip = 33,
# file = "http://www.stat.purdue.edu/~wsc/visualizing.datatables/carbon.dioxide")
# carbon <- cbind(carbon.a, carbon.b)</pre>
# carbon <- t(carbon)
# carbon <- melt(carbon)</pre>
# carbon.ts <- ts(carbon$value, frequency = 12, start = c(1959, 1))
# decompose time series into long term trend, seasonal trend and remaining variation
\# decomp <- stl(carbon.ts, 25, 1)
# tidy data for qqplot2
# carbon <- decomp$time.series</pre>
# years <- floor(time(carbon))</pre>
# months <- rep(1:12, 32)
# carbon <- as.data.frame(carbon)</pre>
# carbon$year <- as.numeric(years)</pre>
# carbon$month <- months
NASA Atmospheric data
# nasa data set loaded with qqsubplot package
# changing temperature units to fahrenheit
# nasa\$fahrenheit <- nasa\$surftemp * 9 / 5 - 459.67
# nasa$atmos.fahrenheit <- nasa$temperature * 9 / 5 - 459.67</pre>
# subset of data centered on Mexico for Figure 4
# subnasa <- subset(nasa, long < -80 & lat > 10 & long > -107)
```

```
###
               diamonds data
# diamonds data set loaded with ggplot2 package
ordinary.diamonds <- subset(diamonds, price <= 5000 & price >= 700 &
 price < (7000 * carat - 1500) & price > (6000 * carat - 2000))
wikileaks Afghan War Diary
# casualties data set loaded with qqsubplot and used as is in figure 2
# regional casualty data included as a supplemental file to paper
# how about casualties over time in different parts of the country?
load("casualties-by-region.RData")
United States Temperature Data
load("seasons.RData")
# trimming data
set.seed(1)
stns <- sample(seasons$stn, 500)
seasons <- seasons[seasons$stn %in% stns, ]</pre>
```

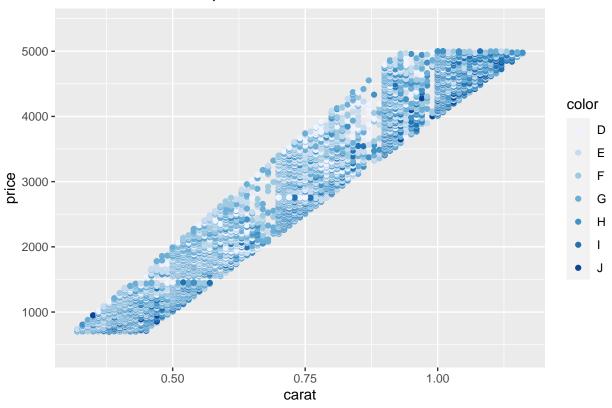
The Struggle part 2: Electric Boogaloo

Poniżej nasza próba załadowania tych figur które się da.

```
# 1-figures.r
# make the figures in the paper
library(RColorBrewer)
library(ggplot2)
#library(ggsubplot)
#source("0-clean.r")
###
                    Figure 1
# Figure 1.a. Mauna Lau CO2 seasonal components (also re-used as Figure 5.a)
ggplot(carbon) +
 geom_subplot(aes(month, mean(seasonal, na.rm = TRUE), group = month,
   subplot = geom_line(aes(year, seasonal - mean(seasonal)))),
   ref = ref_hline(fill = "grey65")) +
 scale_y_continuous(expression(paste("Seasonal component of trend in ", CO[2],
  " (ppm)"))) +
```

```
scale_x_continuous("month", breaks = c(1:12),
   labels = c("J", "F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")) +
  ggtitle("Seasonal frequency components of Mauna Lau carbon \ndioxide time series between 1959 and 199
## Error in ggplot(carbon): object 'carbon' not found
# Figure 1.b. temperature fluctuations
ggplot(nasa) +
  map americas +
  geom_subplot(aes(long, lat, group = id,
    subplot = geom_star(aes(r = fahrenheit, angle = date,
     fill = mean(fahrenheit)), r.zero = FALSE))) +
  scale_fill_gradient("Average\nTemp (F) ",
   breaks = c(55,60,65,70,75,80), guide = "legend") +
  ggtitle("Surface temperature fluctuations 1995 - 2001")
## Error in ggplot(nasa): object 'nasa' not found
# Figure 1.c. binned diamonds data
ggplot(ordinary.diamonds) +
  geom_subplot2d(aes(carat, price,
    subplot = geom_bar(aes(color, fill = color), position = "dodge",
    color = "black", size = 1/10)), bins = c(10, 14), y_scale = free,
   height.adjust = 0.8, width.adjust = 0.8,
   ref = ref_box(aes(color = length(color)))) +
  scale_color_gradient("Total\ncount", low = "grey70", high = "black", guide = "legend") +
  scale_fill_brewer(palette = "Blues") +
  ggtitle("Diamonds, carat vs. price") +
  coord_cartesian(ylim = c(400, 5400))
## Error in geom_subplot2d(aes(carat, price, subplot = geom_bar(aes(color, : could not find function "g
# Figure 1.d. raw diamonds data
ggplot(ordinary.diamonds) +
  geom_point(aes(carat, price, color = color)) +
  scale_color_brewer(palette = "Blues") +
  ggtitle("Diamonds, carat vs. price") +
  coord_cartesian(ylim = c(400, 5400))
```

Diamonds, carat vs. price



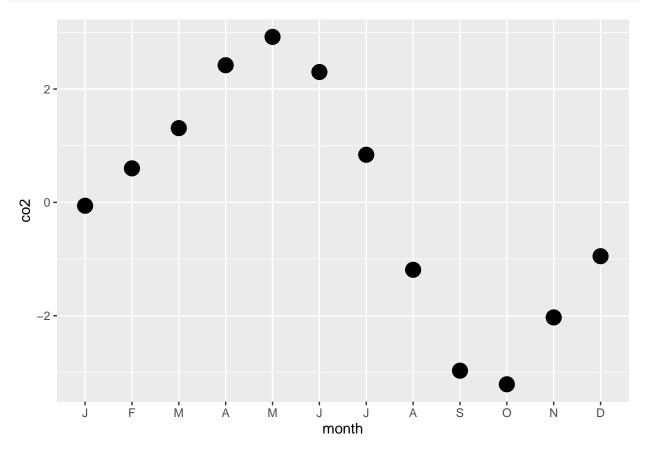
```
Figure 2
# Figure 2.a. raw Afghanistan casualty data
ggplot(casualties) +
 map_afghan +
 geom_point(aes(lon, lat, color = victim), size = 1.75) +
 ggtitle("location of casualties by type") +
 coord_map() +
 scale_colour_manual(values = rev(brewer.pal(5,"Blues"))[1:4])
## Error in ggplot(casualties): object 'casualties' not found
# Figure 2.b. Afghanistan casualty heat map
ggplot(casualties) +
 map_afghan +
 geom_bin2d(aes(lon, lat), bins = 15) +
 ggtitle("number of casualties by location") +
 scale_fill_continuous(guide = guide_legend()) +
 coord_map()
## Error in ggplot(casualties): object 'casualties' not found
# Figure 2.c. Afghanistan casualty embedded bar graphs (marginal distributions)
ggplot(casualties) +
 map_afghan +
 geom_subplot2d(aes(lon, lat,
```

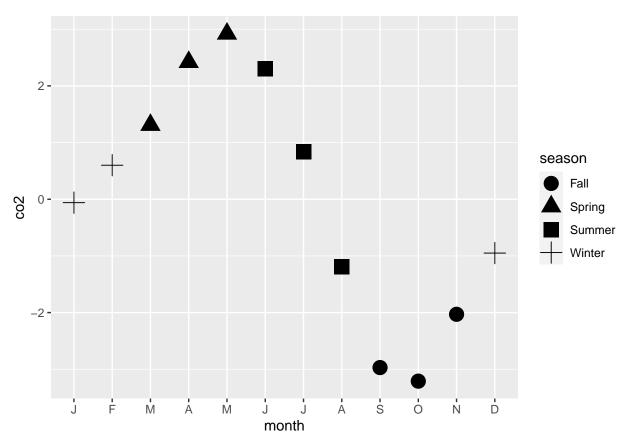
```
subplot = geom_bar(aes(victim, ...count.., fill = victim),
     color = rev(brewer.pal(5, "Blues"))[1], size = 1/4)), bins = c(15,12),
     ref = NULL, width = rel(0.8), height = rel(1)) +
 ggtitle("casualty type by location\n(Marginal distribution)") +
 coord_map() +
 scale_fill_manual(values = rev(brewer.pal(5, "Blues"))[c(1,4,2,3)])
## Error in ggplot(casualties): object 'casualties' not found
# Figure 2.d. Afghanistan casualty embedded bar graphs (conditional distributions)
ggplot(casualties) +
 map afghan +
 geom_subplot2d(aes(lon, lat,
   subplot = geom_bar(aes(victim, ...count.., fill = victim),
     color = rev(brewer.pal(5,"Blues"))[1], size = 1/4)), bins = c(15,12),
     ref = ref_box(fill = NA, color = rev(brewer.pal(5, "Blues"))[1]), width = rel(0.7), height = rel(0
 ggtitle("casualty type by location\n(Conditional distribution)") +
 coord map() +
 scale_fill_manual(values = rev(brewer.pal(5,"Blues"))[c(1,4,2,3)])
## Error in ggplot(casualties): object 'casualties' not found
Figure 3
# Figure 3.a. Temperature data as embedded scatterplots
ggplot(subnasa) +
 map_north +
 geom_subplot(aes(long[1], lat[1], group = id,
   subplot = geom_point(aes(x = surftemp, y = temperature), size = 1/2)),
     width = 3, height = 3) +
 ggtitle("Surface temp vs. atmospheric temp") +
 coord_map()
## Error in ggplot(subnasa): object 'subnasa' not found
# Figure 3.b. Temperature data as embedded line graphs
# notice free x and y scales
ggplot(subnasa) +
 map_north +
 geom_subplot(aes(long[1], lat[1], group = id,
   subplot = geom_smooth(aes(x = surftemp, y = temperature), se = F)),
   x_scale = free, y_scale = free) +
 ggtitle("Surface temp vs. atmospheric temp") +
 coord_map()
## Error in ggplot(subnasa): object 'subnasa' not found
# Figure 3.c. Temperature data as heat map
ggplot(subnasa) +
 lines_north +
 geom_subplot(aes(long[1], lat[1], group = id,
   subplot = geom_blank(aes(long, lat))),
   width = rel(1), height = rel(1),
   ref = ref_box(aes(fill = mean(fahrenheit)), alpha = 0.8, color = NA)) +
```

```
ggtitle("Average temperature (F)") +
scale_fill_continuous("Average\nTemp (F)", guide = "legend", breaks = c(55, 60, 65, 70, 75, 80)) +
coord_map()
```

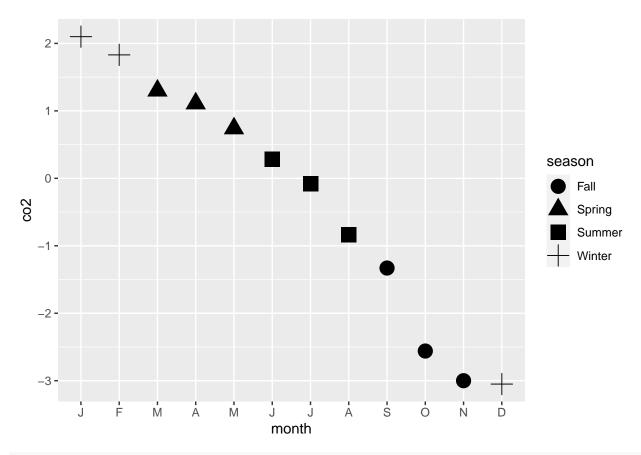
Error in ggplot(subnasa): object 'subnasa' not found

```
###
                     Figure 4
example_data <- structure(list(</pre>
 month = factor(1:12, labels = c("Jan", "Feb", "March", "April",
    "May", "June", "July", "Aug", "Sept", "Oct", "Nov", "Dec")),
 co2 = c(-0.06, 0.6, 1.31, 2.42, 2.92, 2.3, 0.84, -1.19, -2.97,
         -3.21, -2.03, -0.95),
 season = c("Winter", "Winter", "Spring", "Spring", "Spring",
          "Summer", "Summer", "Fall", "Fall", "Fall",
          "Winter")),
 .Names = c("month", "co2", "season"),
 row.names = c(NA, -12L), class = "data.frame")
ggplot(example_data) +
 geom_point(mapping = aes(x = month, y = co2), size = 5) +
 scale_x_discrete(labels = c("J", "F", "M", "A", "M", "J", "J",
                         "A", "S", "O", "N", "D"))
```





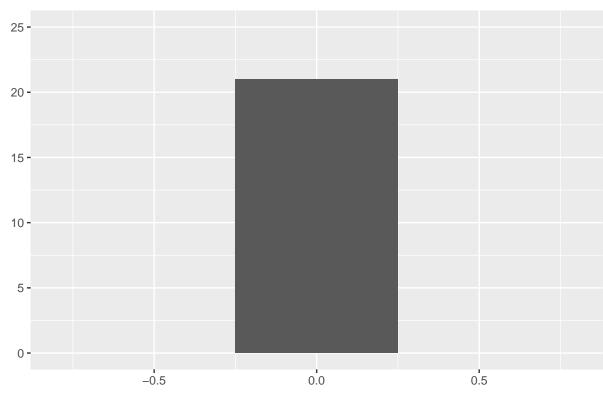
```
Figure 6
other_data <- structure(list(</pre>
 month = factor(1:12, labels = c("Jan", "Feb", "March", "April",
     "May", "June", "July", "Aug", "Sept", "Oct", "Nov", "Dec")),
 co2 = c(2.1, 1.83, 1.3, 1.11, 0.74, 0.28, -0.08, -0.84, -1.33, -2.56,
         -3.00, -3.05),
 season = c("Winter", "Winter", "Spring", "Spring", "Spring",
          "Summer", "Summer", "Fall", "Fall", "Fall",
          "Winter")),
 .Names = c("month", "co2", "season"),
 row.names = c(NA, -12L), class = "data.frame")
ggplot(other_data) +
 geom_point(mapping = aes(x = month, y = co2, shape = season), size = 5) +
 scale_x_discrete(labels = c("J", "F", "M", "A", "M", "J", "J",
                         "A", "S", "O", "N", "D"))
```



```
## Error in ggplot(carbon[carbon$month == 3, ]): object 'carbon' not found
# Figure 7.c. Carbon Dioxide seasonal components embedded plot as scatterplot
ggplot(carbon) +
  geom_subplot(aes(month, mean(seasonal, na.rm = TRUE), group = month,
    subplot = geom_blank(aes(x = month, y = seasonal))),
    ref = ref_box(fill = "grey10", color = NA)) +
  scale_y_continuous("Seasonal component") +
  scale_x_continuous(breaks = c(1:12),
    labels = c("J","F", "M", "A", "M", "J", "J", "A", "S", "O", "N", "D")) +
  ggtitle("Average seasonsal\ncomponent by month")
```

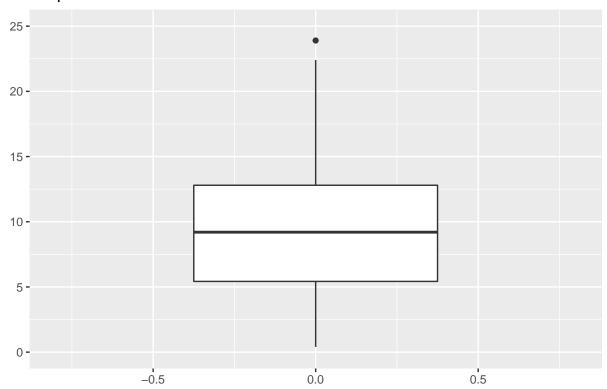
Error in ggplot(carbon): object 'carbon' not found

bar



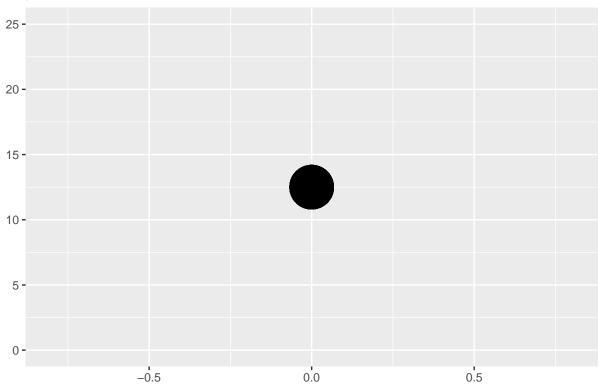
```
# Figure 8.b boxplot geom
ggplot(mtcars) + geom_boxplot(aes(0, mpg - 10)) +
    scale_x_continuous("") +
    scale_y_continuous("") +
    ggtitle("boxplot") +
    coord_cartesian(xlim = c(-0.8, 0.8), ylim = c(0, 25))
```

boxplot



```
# Figure 8.c point geom
ggplot(mtcars) + geom_point(aes(0, 12.5), size = 15) +
    scale_x_continuous("") +
    scale_y_continuous("") +
    ggtitle("point") +
    coord_cartesian(xlim = c(-0.8, 0.8), ylim = c(0, 25))
```





```
Figure 8
# Figure 9.a scatterplot
ggplot(nasa) +
geom_point(aes(x = surftemp, y = temperature)) +
ggtitle("Surface temperature vs. atmospheric temperature (all observations)")
## Error in ggplot(nasa): object 'nasa' not found
# Figure 9.b scatterplot subplots
ggplot(nasa) +
  map_americas +
  geom_subplot(aes(x = long,
                y = lat,
                group = id,
                subplot = geom_point(aes(x = surftemp,
                                     y = temperature),
                                  size = .5, width = 3, height = 3)) +
 ggtitle("Surface temperature vs. atmospheric temperature (by location)")
## Error in ggplot(nasa): object 'nasa' not found
# Figure 9.c star subplots, same as
# Figure 1.b. temperature fluctuations
ggplot(nasa) +
 map_americas +
```

Error in ggplot(nasa): object 'nasa' not found

ylab("Minimum temperature") +
xlab("Maximum temperature")

breaks = c(55,60,65,70,75,80), guide = "legend") +

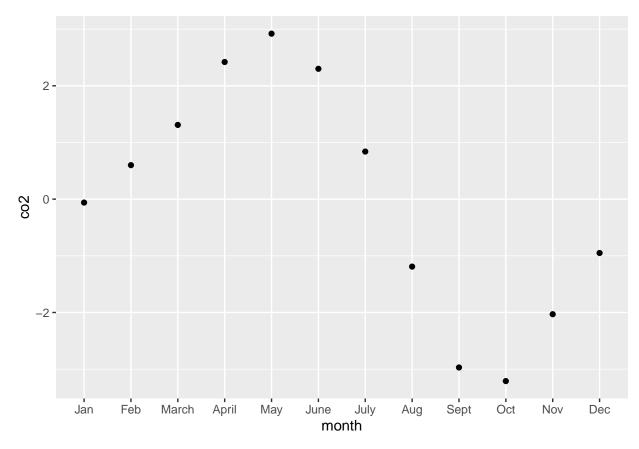
ggtitle("Temperature fluctuations arranged by minimum and maximums.") +

Jak widać ładuje się 7 figur na 23, czyli trochę mniej niż 1/3. Z czego te 7 ładujących się wykresów są to najprostsze wykresy które nie używają pakietu który jest opisywany.

The Struggle part 2.2: More Electric Boogaloo

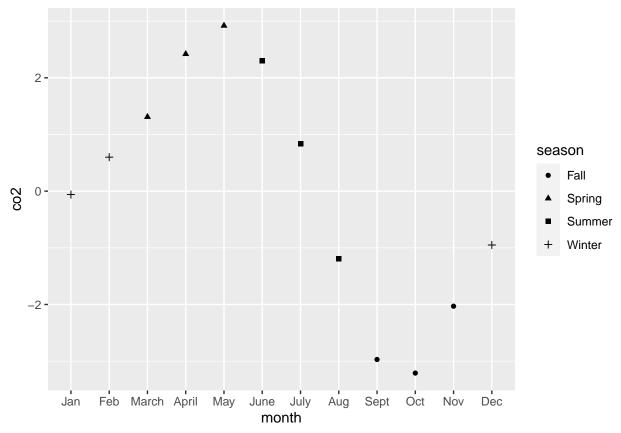
Poniżej znajdują się jeszcze czyste przykłady których tekst jest zamieszczony w samym artykule. Jak przykłady powyżej, ładują się tylko figury niekorzystające z opisywanego pakietu. Tym razem jest to 2/7 figur.

```
# 2-examples.r
# the example code included in the paper and appendix
## Section 4, Figure 4
# Not shown:
example_data <- structure(list(month = structure(1:12,</pre>
  .Label = c("Jan", "Feb", "March", "April", "May", "June", "July",
             "Aug", "Sept", "Oct", "Nov", "Dec"), class = "factor"),
  co2 = c(-0.06, 0.6, 1.31, 2.42, 2.92, 2.3, 0.84, -1.19, -2.97,
          -3.21, -2.03, -0.95),
  season = c("Winter", "Winter", "Spring", "Spring", "Spring",
             "Summer", "Summer", "Fall", "Fall", "Fall",
             "Winter")),
  .Names = c("month", "co2", "season"),
  row.names = c(NA, -12L), class = "data.frame")
# Shown:
library(ggplot2)
ggplot(example_data) +
  geom_point(mapping = aes(x = month, y = co2))
```



```
## Section 4, Figure 5

ggplot(example_data) +
  geom_point(mapping = aes(x = month, y = co2, shape = season))
```



```
## Section 4, Figure 7
library(ggsubplot)
## Error in library(ggsubplot): there is no package called 'ggsubplot'
ggplot(carbon) +
  geom_subplot(aes(x = month,
  y = mean(seasonal, na.rm = TRUE),
   group = month,
   subplot = geom_line(aes(x = year,
     y = seasonal - mean(seasonal)))))
## Error in ggplot(carbon): object 'carbon' not found
## Section 4, Figure 9
ggplot(nasa) +
  geom_point(aes(x = surftemp, y = temperature))
## Error in ggplot(nasa): object 'nasa' not found
ggplot(nasa) +
  map_americas +
  geom_subplot(aes(x = long,
    y = lat,
    group = id,
    subplot = geom_point(aes(x = surftemp,
```

```
y = temperature))))
## Error in ggplot(nasa): object 'nasa' not found
ggplot(nasa) +
  map_americas +
  geom_subplot(aes(x = long,
    y = lat,
    group = id,
    subplot = geom_star(aes(r = fahrenheit,
      angle = date,
      fill = mean(fahrenheit)),
      r.zero = FALSE)))
## Error in ggplot(nasa): object 'nasa' not found
ggplot(nasa) +
  geom_subplot(aes(x = max(surftemp),
    y = min(surftemp),
    group = id,
    subplot = geom_star(aes(r = surftemp,
      angle = date,
      fill = mean(fahrenheit)),
      r.zero = FALSE)))
```

Error in ggplot(nasa): object 'nasa' not found

The End of Evangelion Struggle

Podsumowując: artykuł okazał się niereprodukowalny. Pakiet jest przestarzały i porzucony przez autorów. Część zbiorów danych jest też niemożliwa do pobrania. Pakiet miał potencjał ale nie został on wykorzystany.