Analiza imprez masowych na terenie Polski w poszczególnych powiatach

Jan Kowalski, Michał Nowak

20 czerwca 2019

## Wczytywanie i przygotowanie danych

Dane potrzebne do analizy zostały pobrane z Banku Danych Lokalnych. Dane przedstawiają ilość zorganizowanych imprez masowych, liczbę ludności oraz dochody budżetowe poszczególnych powiatów. Na ogólną sumę imprez masowych składają się imprezy artystyczno-rozrywkowe (m.in. koncerty, festiwale, rekonstrukcje historyczne), interdyscyplinarne oraz sportowe. Analiza została wykonana na podstawie danych z roku 2017.

library(dplyr)  
library(sf)  
library(raster)  
library(tidyverse)  
library(spData)  
library(spDataLarge)  
library(tmap) # for static and interactive maps  
library(leaflet) # for interactive maps  
library(mapview) # for interactive maps  
library(ggplot2) # tidyverse vis package  
library(shiny) # for web applications  
library(readxl)  
  
powiaty <- st\_read('J:/ADP - projekt01/projekt/data/Powiaty.shp',   
 stringsAsFactors = F)

## Reading layer `Powiaty' from data source `J:\ADP - projekt01\projekt\data\Powiaty.shp' using driver `ESRI Shapefile'  
## Simple feature collection with 380 features and 29 fields  
## geometry type: MULTIPOLYGON  
## dimension: XY  
## bbox: xmin: 14.12288 ymin: 49.00205 xmax: 24.14578 ymax: 54.83642  
## epsg (SRID): NA  
## proj4string: +proj=longlat +ellps=GRS80 +no\_defs

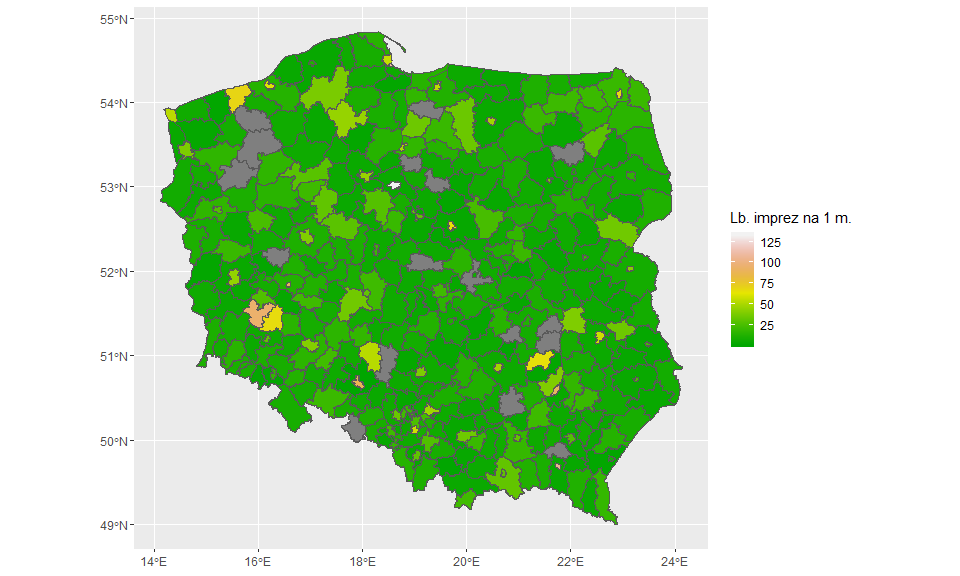
atrybuty <- read\_xlsx('J:/ADP - projekt01/projekt/data/KULT\_3402\_XTAB\_20190602123115.xlsx')  
imprezy <- left\_join(powiaty,atrybuty, by = c('JPT\_KOD\_JE'='Kod'))  
  
ludnosc <- read\_xlsx('J:/ADP - projekt01/projekt/data/LUDNOSC.xlsx')  
imprez\_ludnosc <- left\_join(imprezy, ludnosc, by = c('JPT\_KOD\_JE' = 'Kod'))  
  
finanse <- read\_xlsx('J:/ADP - projekt01/projekt/data/Finanse.xlsx')  
imp\_lud\_fin <- left\_join(imprez\_ludnosc, finanse, by = c('JPT\_KOD\_JE' = 'Kod'))  
  
dane <- imp\_lud\_fin %>%  
 dplyr::select('suma', 'Nazwa.x', 'ludnosc', 'finanse')

## Wizualizacja danych

### Ilość imprez masowych w poszczególnych powiatach

Aby zwizualizować ilość imprez masowych została dodana nowa kolumna “imprezy\_na\_ludnosc”, która jest ilorazem sumy imprez i liczby ludności w danym powiecie.

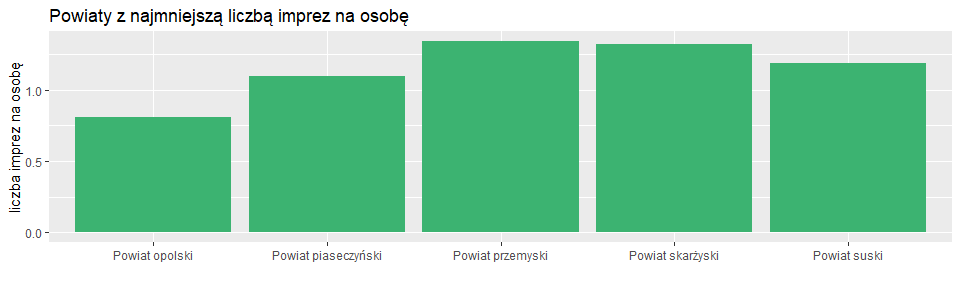
#dodanie kolumny imprezy/ludnosc  
dane$imprezy\_na\_ludnosc <- (as.numeric(dane$suma)/as.numeric(dane$`ludnosc`))\*100000  
  
ggplot()+  
 geom\_sf(data = dane, aes(fill = imprezy\_na\_ludnosc)) +   
 scale\_fill\_gradientn(colors = terrain.colors(20)) +  
 labs(fill = "Lb. imprez na 1 m.")

 Najwięcej imprez masowych w 2017 roku odbyło się w powiecie m. Toruń (ok. 133 na osobę). Kolejne powiaty to: - m. Sopot (112/os.) - m. Krosno (105/os.) - polkowicki (95/os.) Najmniej imprez odbyło się w powiecie opolskim (mniej niż jedno wydarzenie na osobę).

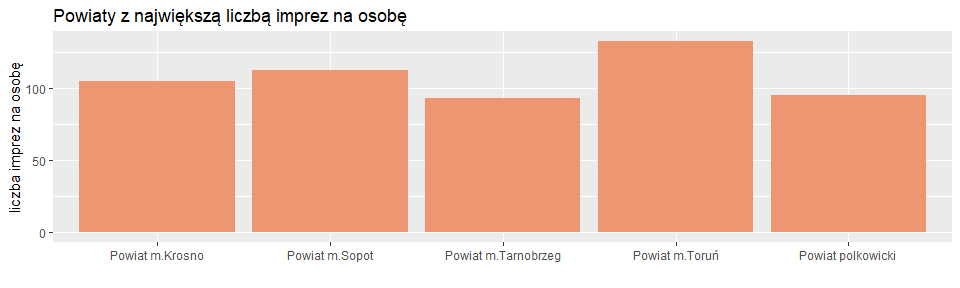
Poniższe wykresy przedstawiają powiaty o największej oraz najmniejszej liczbie imprez masowych przeliczonych na osobę w danym powiecie.

dane2 <- na.omit(dane)  
  
powiaty\_max\_imp <- dane2 %>%   
 dplyr::arrange(desc(dane2$imprezy\_na\_ludnosc)) %>%   
 head(5)   
  
powiaty\_min\_imp <- dane2 %>%  
 dplyr::arrange(desc(dane2$imprezy\_na\_ludnosc)) %>%   
 tail(5)  
  
  
wykres1 <- ggplot(powiaty\_min\_imp,   
 aes(x = powiaty\_min\_imp$Nazwa.x,   
 y = powiaty\_min\_imp$imprezy\_na\_ludnosc)) +   
 geom\_col(fill= 'mediumseagreen') +  
 labs(title = 'Powiaty z najmniejszą liczbą imprez na osobę',   
 x= "",   
 y = "liczba imprez na osobę")  
  
wykres2 <- ggplot(powiaty\_max\_imp,   
 aes(x = powiaty\_max\_imp$Nazwa.x,   
 y = powiaty\_max\_imp$imprezy\_na\_ludnosc)) +   
 geom\_col(fill= 'lightsalmon2') +  
 labs(title = 'Powiaty z największą liczbą imprez na osobę',   
 x= "",   
 y = "liczba imprez na osobę")

wykres1

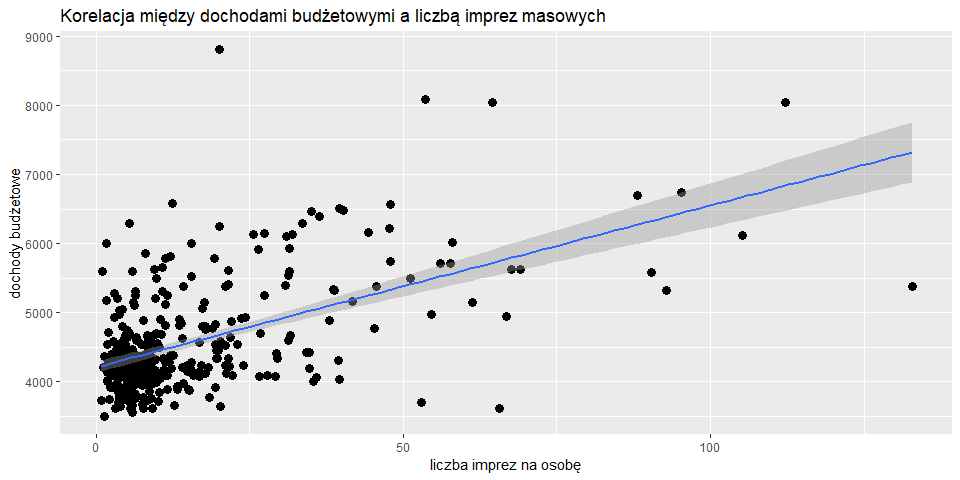


wykres2



Powszechnie wiadomo, że sytuacja finansowa danej jednostki terytorialnej wpływa na ilość zorganizowanych imprez masowych. Jak silna jest ta zależność przedstawia poniższa analiza:

ggplot(dane2,aes(x = imprezy\_na\_ludnosc, y= finanse)) +   
 geom\_point(size = 3) +   
 geom\_smooth(method = lm) +  
 labs(title = "Korelacja między dochodami budżetowymi a liczbą imprez masowych",   
 x= "liczba imprez na osobę",   
 y = "dochody budżetowe")



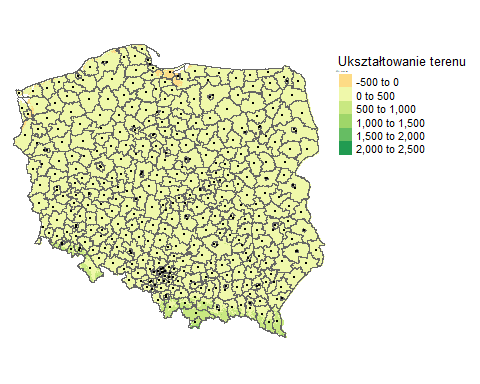
cor.test(dane2$imprezy\_na\_ludnosc, dane2$finanse)

##   
## Pearson's product-moment correlation  
##   
## data: dane2$imprezy\_na\_ludnosc and dane2$finanse  
## t = 12.58, df = 361, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.4762034 0.6197882  
## sample estimates:  
## cor   
## 0.5520755

W celu zbadania zależności między dochodami budżetowymi a liczbą imprez masowych został wykorzystany współczynnik Pearsona. Jego wartość wynosi ok. 0,55, a więc zależność między zmiennymi mieści się w teoretycznym przedziale 0,5-0,7 i jest silna. Znak przy współczynniku Pearsona jest dodatni, co oznacza, że im wyższe dochody budżetowe, tym wyższa liczba imprez masowych. Ponadto, wyliczona p-wartość (p-value) jest mniejsza niż 0,05, co oznacza, iż zmienna objaśniająca (dochody budżetowe) jest istotna.

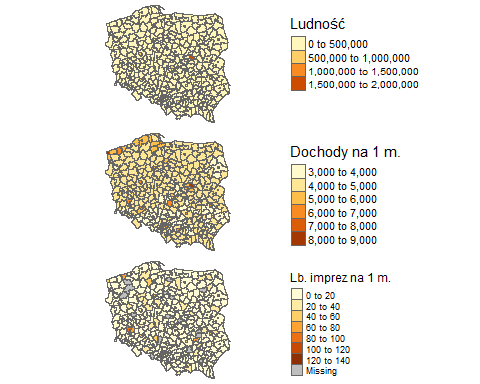
Dopełnieniem analizy jest mapa Polski z podziałem na powiaty wraz z oznaczonymi centroidami. Na mapie Polski zostało również zaznaczone ukształtowanie terenu.

#t\_map dla powiaty + punkty centroidy  
library(tmap) # for static and interactive maps  
dem <- getData('alt', country = "PL")  
powiaty2<- transform(powiaty, 4326)  
m1 <- tm\_shape(dem) + tm\_raster()  
m2 <- m1 + tm\_shape(powiaty) + tm\_borders()  
centroidy <- st\_centroid(powiaty2)  
m3 <- m2 + tm\_shape(centroidy) + tm\_dots()  
  
  
tm\_shape(dem) +   
 tm\_raster()+ #warstwa rastra wysokoĹ›ci nad poziomem morza  
 tm\_shape(powiaty) +   
 tm\_borders() +   
 tm\_shape(st\_centroid(powiaty2)) +   
 tm\_layout(legend.outside=TRUE,   
 legend.outside.position = 'right',   
 legend.title.size = 0.1,   
 frame = FALSE) +  
 tm\_legend(title = 'Ukształtowanie terenu')+  
 tm\_dots()



Poniższe kartogramy zostały pokolorowane względem zmiennych wchodzących udział w całej analizie.  
Pierwszy kartogram przedstawia liczbę ludności w poszczególnych powiatach, kolejny wysokość dochodów na osobę, a ostatni to liczbę imprez masowych na jednego mieszkańca powiatu.

map1 = tm\_shape(dane) +   
 tm\_borders() +   
 tm\_fill(col = "ludnosc", title = "Ludność") +   
 tm\_polygons(palette = "BuGn")+  
 tm\_layout(legend.outside=TRUE,   
 legend.height = 0.3,   
 legend.outside.position = "right",   
 frame = FALSE)  
map2 = tm\_shape(dane) +   
 tm\_borders() +   
 tm\_fill(col = "finanse",   
 title = "Dochody na 1 m.") +  
 tm\_layout(legend.outside=TRUE,   
 legend.height = 0.3,   
 frame = FALSE)  
map3 = tm\_shape(dane) +   
 tm\_borders() +   
 tm\_fill(col = "imprezy\_na\_ludnosc",   
 title = "Lb. imprez na 1 m.") +  
 tm\_layout(legend.outside=TRUE,   
 legend.height = 0.3,   
 frame = FALSE)  
  
  
  
#tytuł map  
legend\_title = expression("Ludnosc")  
map\_nza = map1 +  
 tm\_fill(title = legend\_title)  
  
  
tmap.arrange <- tmap\_arrange(map1, map2, map3, ncol = 1)  
tmap.arrange



# Statystyki strefowe dla Polski

## Pobranie danych oraz podstawowe statystyki

W celu pobraniu informacji dotyczących poligonu - Polski, skorzystano z funkcji getData() - wybierając kraj, który jest przedmiotem analizy. Otrzymano podstawowe informacje określające ukształtowanie terenu Polski.

#Polska statystyki strefowe   
DEM.Poland <- getData('alt', country = 'pol', mask = TRUE)  
summary(DEM.Poland[])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -11.0 99.0 142.0 168.2 197.0 2079.0 291245

length(DEM.Poland)

## [1] 881280

## Podstawowe statystyki dla poszczególnych województw w Polsce oraz ich wizualizacja

wojewodztwa <- st\_read('J:/ADP - projekt01/projekt/data/Województwa.shp',   
 stringsAsFactors = F)

## Reading layer `Wojewodztwa' from data source `J:\ADP - projekt01\projekt\data\WojewĂłdztwa.shp' using driver `ESRI Shapefile'  
## Simple feature collection with 16 features and 29 fields  
## geometry type: MULTIPOLYGON  
## dimension: XY  
## bbox: xmin: 14.12288 ymin: 49.00205 xmax: 24.14578 ymax: 54.83642  
## epsg (SRID): NA  
## proj4string: +proj=longlat +ellps=GRS80 +no\_defs

wojewodztwa$JPT\_KOD\_JE <- as.numeric(wojewodztwa$JPT\_KOD\_JE)  
wojewodztwa2 <- st\_transform(wojewodztwa, 4326)   
wojewodztwa.ra <- rasterize(wojewodztwa2,   
 DEM.Poland,   
 field = 'JPT\_KOD\_JE',  
 fun = min)  
  
  
  
wojewodztwa.zone <- zonal(DEM.Poland, wojewodztwa.ra, fun = mean)  
wojewodztwa3 <- left\_join(wojewodztwa2, as.data.frame(wojewodztwa.zone), by=c('JPT\_KOD\_JE'='zone'))  
  
library(ggspatial)  
ggplot()+geom\_sf(data = wojewodztwa3, aes(fill=value)) +  
 annotation\_scale(location = "bl", width\_hint = 0.3) +  
 scale\_fill\_gradientn(colours = terrain.colors(20))

