Electoral Geography: Mapping Applications

Rodrigo Rodrigues-Silveira

March 27th, 2017

The data, the code and examples are available at: https://github.com/rodrodr/Spatial_Analysis

Mapping Applications

This document explores some mapping applications in R using electoral geography as a core example. These maps include some useful functions to compare and understand voting behavior. Nonetheless, they also serve to many other situations or research fields. These functions were created to rule out repetitive tasks when analyzing geographic data in R. They automatize routine tasks into simpler features tailored for social sciences mapping applications.

The first step here is to load all functions and treat the example data employed in this brief document. We will use Brazilian electoral data at the municipal level for demonstrating the usefulness and functionality of each R function. The following code loads the source and prepares the data required to perform the examples:

```
source("d:/Dropbox/Stats/Code/Mapping_Applications/Mapping_Applications.R")



library(sp)
library(maps)
library(maptools)

e <- readShapeSpatial("d:/Dropbox/Stats/Code/Spatial_Utilities_Data/Estados.shp")
a <- readShapeSpatial("d:/Dropbox/Stats/Code/Spatial_Utilities_Data/Aerodromos.shp")
b <- readShapeSpatial("d:/Dropbox/Stats/Code/Spatial_Utilities_Data/mapa_base.shp")

a$pavement <- "other"
a$pavement[a$PAVIMENTO=="terra"] <- "dirt runway"
a$pavement[a$PAVIMENTO%in%c("asfalto ou concreto Asfál","concreto")] <- "paved"


uf <- c("SP","MG","RJ","BA","RS","PR","PE","CE","PA","MA","SC","GO","AM","PB",
        "ES","RN","AL","MT","PI","DF","MS","SE","RO","TO","AC","AP","RR")
pop <- c(44.8,21,16.6,15.3,11.3,11.2,9.4,9,8.3,7,6.9,6.7,4,4,4,3.5,3.4,3.3,3.2,
         3.2,2.7,2.3,1.8,1.5,0.8,0.8,0.5)

u <- data.frame(UF=uf, Population=pop)

e <- merge(e,u, by="UF")


### loads the data
load("d:/Dropbox/Stats/Code/Spatial_Utilities_Data/BR_Pres_vote.RData")

bl <- readShapeSpatial(
  "d:/Dropbox/Stats/Code/Spatial_Utilities_Data/Brazilian_Localities.shp")
bl <- bl[,c("GEOCODIG_M", "LONG","LAT")]
```

```r
## Merges electoral data to the geographical locations of Brazilian municipalities
br <- br[order(br$year),]
br <- merge(br, bl, by="GEOCODIG_M", all.x=T)
br <- br[! is.na(br$LONG),]

## Converts the new object into a SpatialPointsDataFrame
coordinates(br) <- ~LONG+LAT


# Converts time series data into a time slices format
year <- seq(1998,2014,4)

dt <- data.frame(br)

for (i in 1:length(year)){

  dta <- dt[dt$year==year[i] & dt$party==13,c("GEOCODIG_M","prop_votes","qt_votes")]
  dtb <- dt[dt$year==year[i] & dt$party==45,c("GEOCODIG_M","prop_votes","qt_votes")]

  names(dta)[2] <- paste0("p_pt", year[i])
  names(dtb)[2] <- paste0("p_psdb", year[i])
  names(dta)[3] <- paste0("qt_pt", year[i])
  names(dtb)[3] <- paste0("qt_psdb", year[i])

  b <- merge(b, dta, by="GEOCODIG_M", all.X=T)
  b <- merge(b, dtb, by="GEOCODIG_M", all.X=T)

}
```

Electoral Competition Map

The first function we explore here is the electoral competition map. Morrill, Knopp, and Brown originally developed this representation in their article "Anomalies in red and blue: Exceptionalism in American electoral geography" (2007) for representing political competition between Democrats and Republicans in te US. This choropleth map compares the performance of two political parties over two elections and converts the results in a diverging color scheme from the best to the worst results for each party. In general terms, it corresponds to an electoral volatility map, once it can capture six possible outcomes of political competition between two parties during two elections.

The categories are the following:

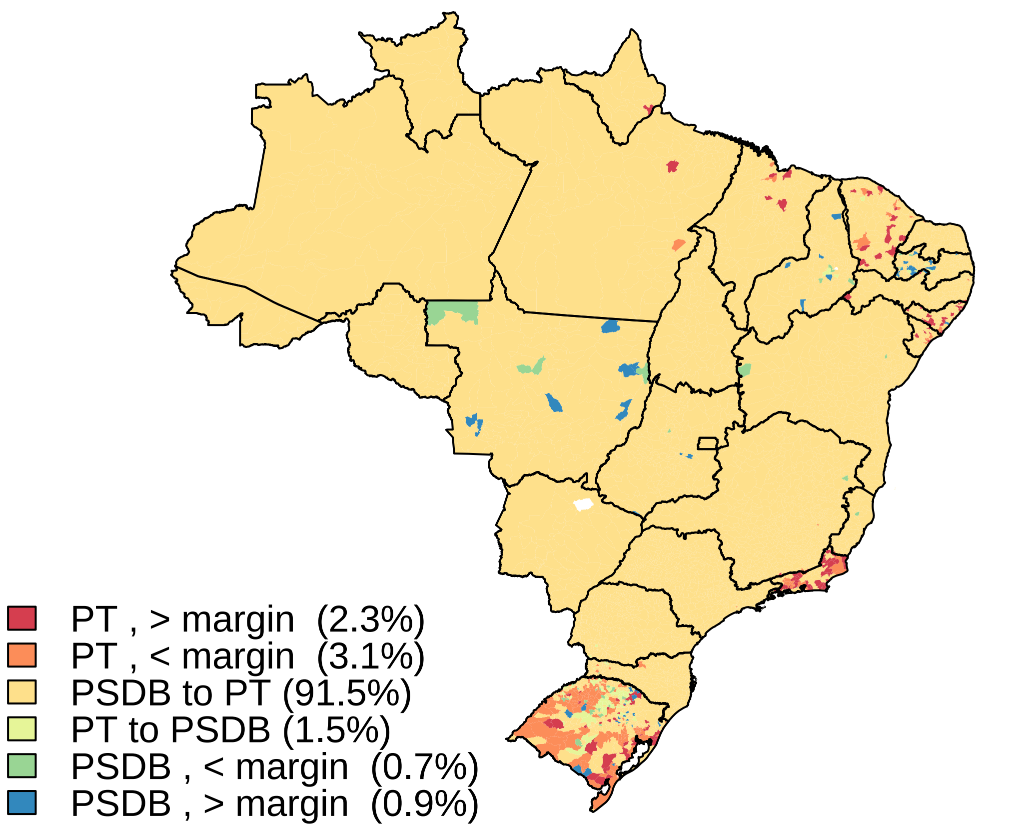
1. **Party A wins with increased margin** – Party A wins in both elections, with a greater margin in the second election. It means that Party A enhanced its performance in these locations over time. In concrete terms, these results mean that Party A reinforced its dominance over these areas. This scenario is the best considering both elections.
2. **Party A with reduced margin** – Party A wins in both elections, but it loses votes in the second one. This result can mean two things. First, Party B is challenging Party A's dominance in this area or, second, the overall competition for the electoral control is higher (particularly true in multiparty systems).
3. **Party A conquers space from Party B** – Party A lost the first election but won the second. It means that it took over locations previously belonging to Party B. When we consider just the last election, this is the best outcome.

4. **Party B conquers space from Party A** – From this category on, the interpretation of the results is the reverse. Party B lost the first election but won the second. It means that it took over locations previously belonging to Party A. When we consider just the last election, this is the best outcome.

5. **Party B with reduced margin** – Party B wins in both elections, but it loses votes in the second one. This result can mean two things. First, Party A is challenging Party B's dominance in this area or, second, the overall competition for the electoral control is higher (particularly true in multiparty systems).
6. **Party B wins with increased margin** – Party B wins in both elections, with a greater margin in the second election. It means that Party B enhanced its performance in these locations over time. In concrete terms, these results mean that Party B reinforced its dominance over these areas. This scenario is the best considering both elections.

In the next map, we compare the electoral results between 1998 and 2002 in Brazilian municipalities for the two main parties: PSDB and PT.

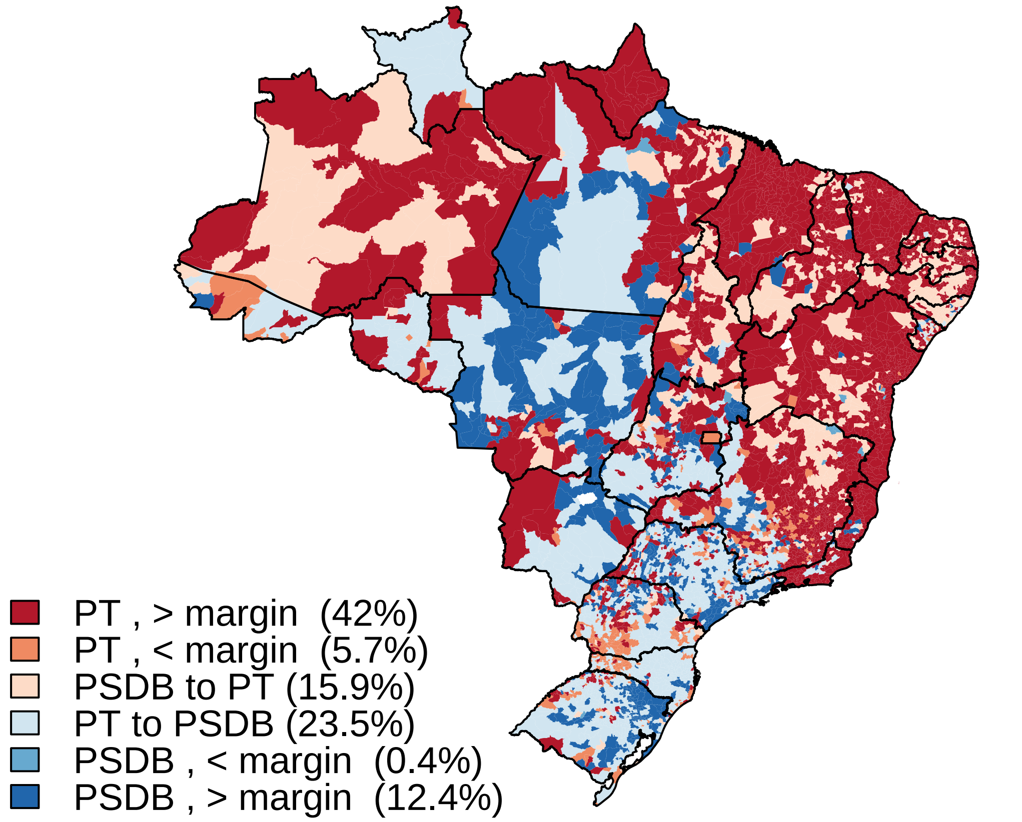
```
# Compares PT and PSDB vote in presidential elections between 1998 and 2002
compmap(b, b$p_pt1998, b$p_psdb1998, b$p_pt2002, b$p_psdb2002, border = "transparent",
        frame = e, leg = c("PT","PSDB"), lang = "en")
```



The results represent a presidential change in a nationalized party system. In this election, Luis Inácio Lula da Silva (PT) won the presidency against the incumbent candidate José Serra (PSDB). The predominance of the color beige reveals a pattern of change. In 1998 elections, Fernando Henrique Cardoso (PSDB) won the reelection with wide margins against the PT. In 2002, Lula was the victorious challenger, presenting higher percentages of the vote (compared to those of the PSDB) in almost the entire country. The most salient aspect we want to highlight is there is no regional pattern of concentration of votes. This spatial pattern will be strikingly different from the next elections.

The following map compares the two parties between 2002 and 2006. The results are clearly distinct from those of the previous map:

```r
# The same for 2002 and 2006
compmap(b, b$p_pt2002, b$p_pt2006, b$p_psdb2002, b$p_psdb2006, border = "transparent",
        frame = e, leg = c("PT","PSDB"), palette = "RdBu", lang = "en")
```



The nationalized behavior observed in the first map (1998-2002) transforms into a regionalized voting pattern. Although Lula won the reelection, the PT lost space for the PSDB in the Center-South region while expanded its presence in the North and Northeast. The nationalization of the presidential vote, valid until 2002, disappears for a regionalized division of party support. Some studies, such as those of Soares and Terron (2008), demonstrated that this is strictly tied to the coverage of Bolsa Familia, a Conditional Cash Transfer benefit targeted at low-income families. According to them, the targeting of social policies changed the base for support of the PT and, therefore, transformed the electoral geography of presidential elections in Brazil from 2006 on.

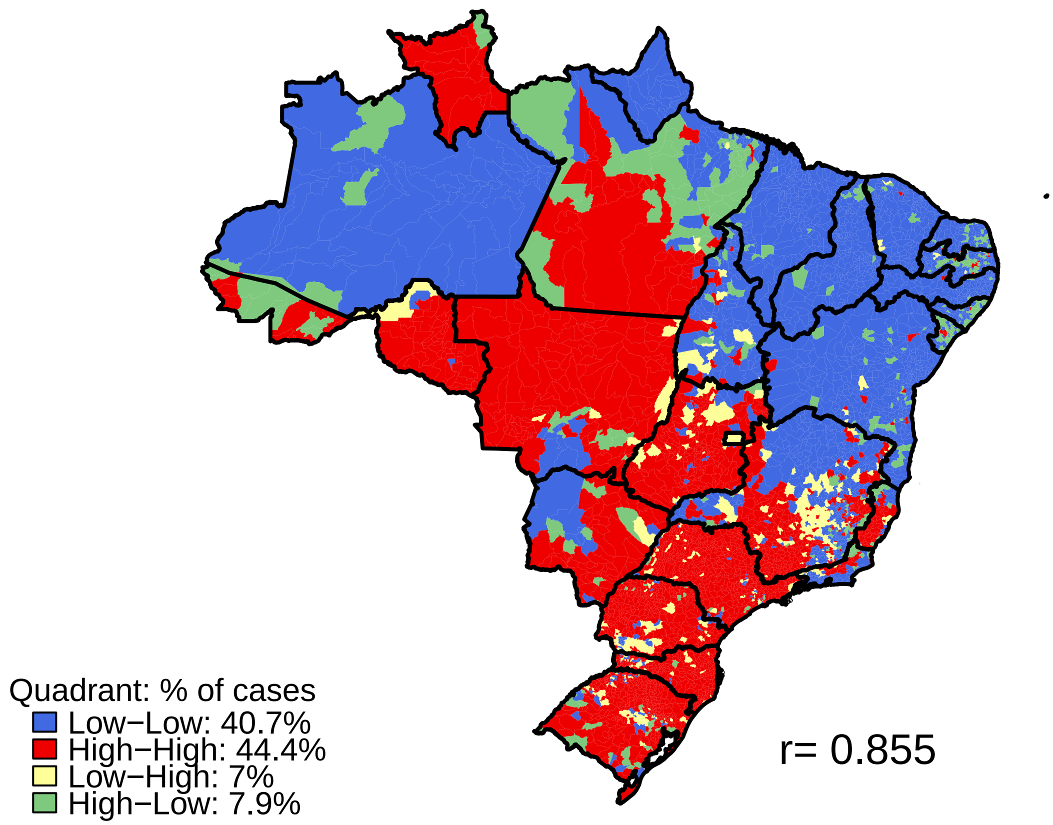
The two maps above are useful tools to visualize this transformation in the electoral geography in Brazil. They make clear that: (a) 2002 was a critical election regarding change; and (b) 2006 can be considered a critical election due to changes introduced in the spatial distribution of political support.

Quadrant Map

Imagine a scatterplot of two standardized variables plotted on a map. This is precisely the purpose of this function. The quadrant map allows for representing correlations and the spatial association of two variables. There are some strong arguments for this kind of representation. On the one hand, high correlations say nothing about the spatial concentration of values. The level of association can be high, but these values can be dispersed all over the territory with no clustering. On the other, and most importantly, correlation coefficients near zero or lower than 0.2 can hide significant patterns of territorial concentration or clusters.

Originally this map was developed by Luc Anselin to represent spatially local Moran's I values. Nonetheless, the same principle applies to any bivariate association. For instance, the next map examines the association between PSDB vote in 2010 and 2014:

```
par(mar=rep(0,4))
res <- qdmap(b, b$p_psdb2010, b$p_psdb2014, border = "transparent", frame=e, ret.col = T)
```



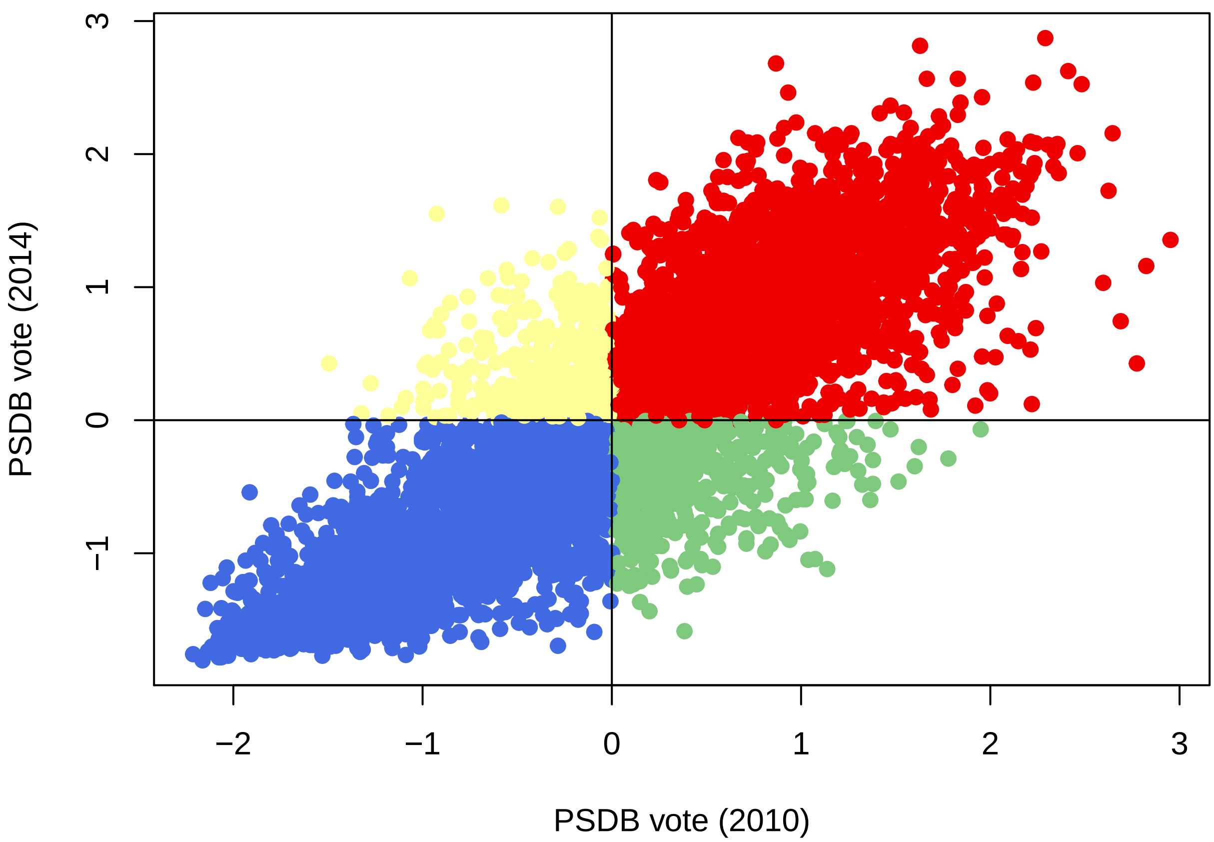
As one can see, there is a regional pattern of clustering of both high and low percentage of votes in both elections. In 2010 and 2014, the PSDB concentrated its votes in the Center-South regions of Brazil, while its performance was worse in parts of the North and almost the entire Northeast. In this case, the correlation is also high, with a Pearson coefficient of 0.855.

There are four colors, each representing one quadrant of a scatterplot where the division line in the X and Y axes is the mean of each variable. Thus, those values that are below the average for both variables are blue. Those cases where both are above the mean are red. High values of X and Low values of Y are painted green. Low values for X and high for Y are yellow. In this particular example, we can read the results as follows:

1. **Blue (Low-Low)-** the percentage of the vote in the PSDB was below the mean for both 2010 and 2014 presidential elections.
2. **Red (High-High)-** the percentage of the vote in the PSDB was above the mean for both 2010 and 2014 presidential elections.
3. **Yellow (Low-High)-** the percentage of the vote in the PSDB was below the mean for the 2010 presidential elections, but above the mean for 2014.
4. **Green (High-Low)-** the percentage of the vote in the PSDB was above the mean for the 2010 presidential elections, but below the mean for 2014.

We can visualize this relationship plotting these colors into a scatterplot:

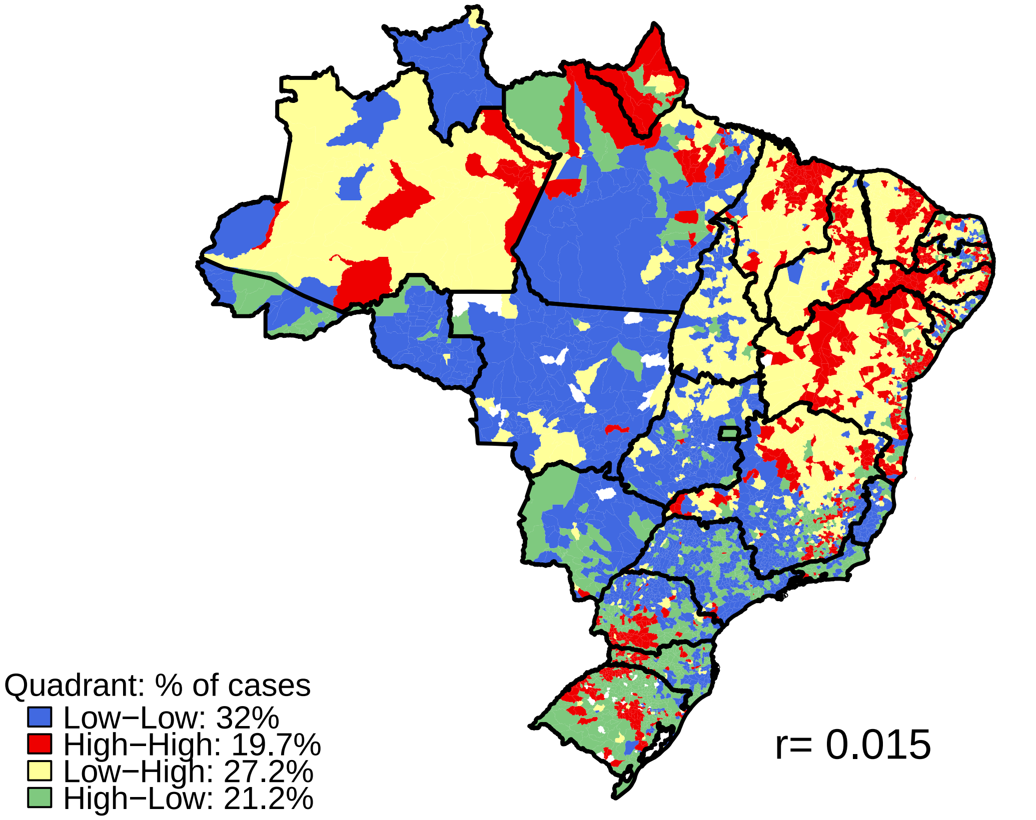
```
par(mar=c(4,4,1,1))
plot(res$zx, res$zy, col=res$colors, pch=19, xlab="PSDB vote (2010)",
     ylab="PSDB vote (2014)")
abline(h=0, v=0)
```



As we should expect, parties' voting patterns tend to be stable over time, so the lagged value for the vote in a given area and election should be highly correlated to the values obtained in the same location in a previous contest. The purpose of this scatterplot is another. We want to demonstrate how a basic statistical procedure, such as a correlation, could be visualized spatially. As we can see, we represent points falling into each quadrant with a color correspondent to their positions according to the mean of both variables. These are the same colors drawn in the map. Most points falls into the Low-Low (blue) and High-High (red) quadrants. This concentration explains the correlation coefficient of 0.855.

Now we repeat the map, but with a different pattern or, more precisely, a change in pattern. We will use presidential vote in the PT in 1998 and 2010. During this period, as we have seen before, PT vote changed radically from a national dispersion to a regional concentration. Differently from the PSDB, this time the correlation is similar to zero (r=0.015), so the question is: is there any spatial pattern hidden behind this low coefficient? The anwer shown in the map below reveals that, even in those cases where there is no correlation, we can observe systematic clustering of low and how electoral performance. It is possible also to spot the regions that are responsible for the most part of the change. In this map, two regions outstand: The extreme south and the Northeast.

```
par(mar=rep(0,4))
qdmap(b, b$p_pt1998, b$p_pt2010, border = "transparent", frame=e)
```

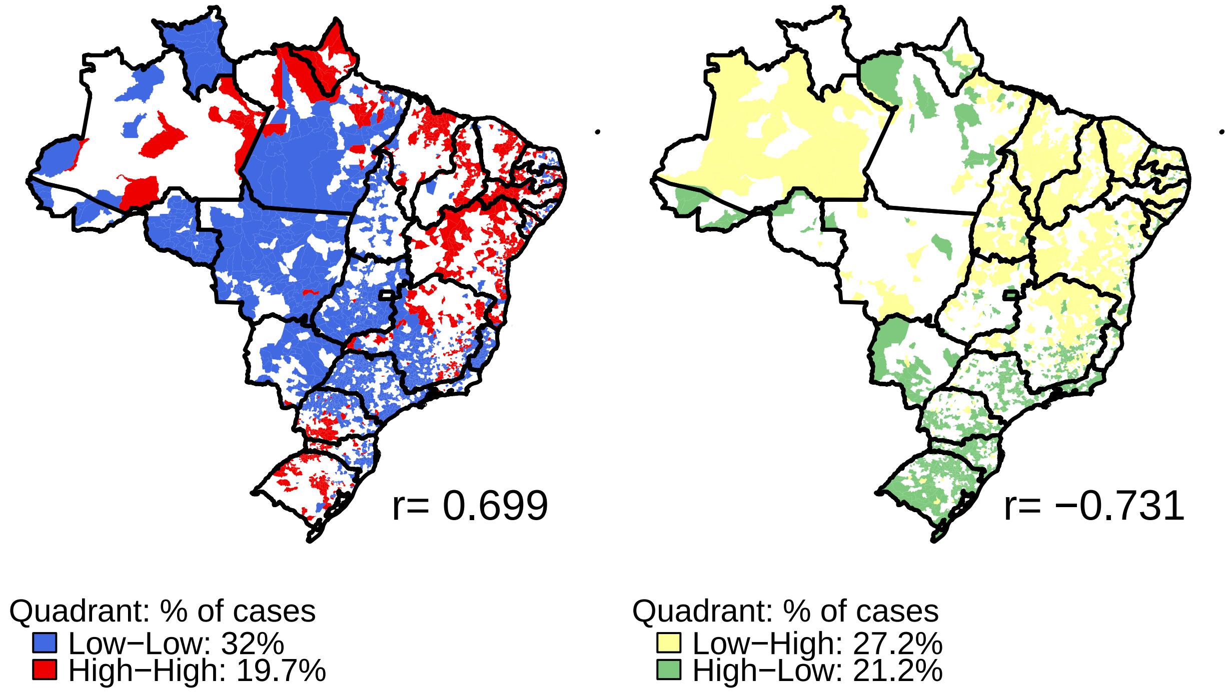


Sometimes, as in this particular case, it is difficult to identify clustering patterns. In these cases, to facilitate the visualization, we can also decompose the map into pairs of association. If we assume a positive correlation, we would like to inspect whether cases are concentrated in blue and red. On the other hand, negative associations would shift our attention towards green and yellow values. This procedure is also very useful when the correlation coefficient is close to zero. Due to phenomena such as spatial heterogeneity, it is possible that a zero coefficient is hiding two opposite correlations that cancel themselves but are coherent spatially.

The map below implements this separation for the PT vote between 1998 and 2010. The first map focuses only in the High-High and Low-Low values and the second in the Low-High and High-Low values. We can identify more clearly the areas of concentration of values for each category. The first map shows stability, i.e., places where the vote was consistent over the period. The second represents change; places where vote is not coherent form one election to another.

As the previous maps, the quadrant map is a visual tool for assessing spatial patterns not easily captured by other statistics or visual techniques. More than that, it builds on existing techniques (correlation and scatterplot) to add the spatial dimension in order to explore geographical patterns and relations. It is useful because it alows us to explore bivariate relations adding to it a third (space) and a fourth, time, when comparing the same phenomenon in different periods or moments.

```
par(mfrow=c(1,2))
par(mar=rep(0,4))
qdmap(b, b$p_pt1998, b$p_pt2010, border = "transparent", frame=e, selqad = 2)
qdmap(b, b$p_pt1998, b$p_pt2010, border = "transparent", frame=e, selqad = 3)
```




```r
par(mfrow=c(1,1))
```

High-Low Map

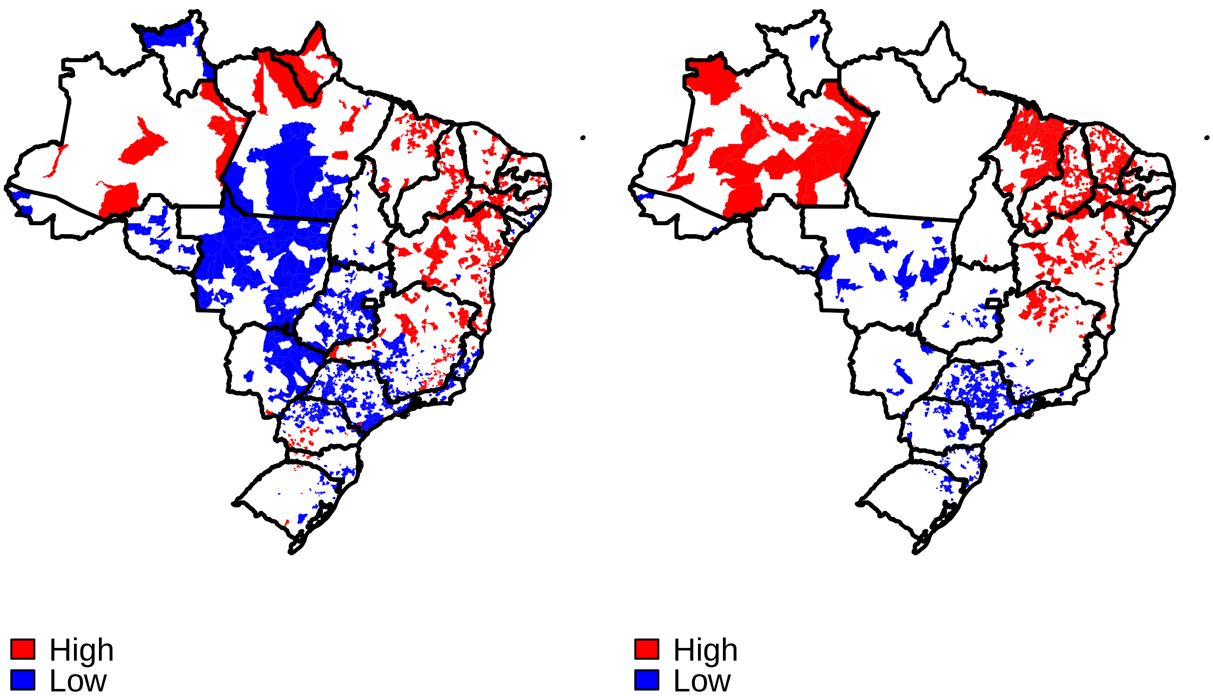
Although the previous map allows us to compare the distribution of two different variables according to their respective means, sometimes we need to inspect the endurance of certain patterns over broader series of time or multiple indicators. One key assumption in electoral geography considers political phenomena as relatively stable in the territory over time. In this regard, we need to know which areas are consistently voting in favor or against a particular political party over the last three, four, five, or even more elections.

The high-Low map accomplishes this task by highlighting those areas that, in a given matrix of event, are consistently above and below the mean in every section or variable. It also provides the possibility of defining the distance of the mean to represent values (in z-score values). The following maps displays those municipalities that voted for the PT both above and below the average consistently in the last five presidential elections.

The first map describes any locality above or below average in each election. The second establish the condition of a z-value higher than one standard deviation in order to represent values. This optional parameter allows us to check not just for consistency, but also for intensity. By this procedure it is possible to answer more precise questions such as: what are the places where PT voting was significantly higher or lower than the average in the last three elections?

```r
par(mfrow=c(1,2))
par(mar=rep(0,4))

# PT - includes the period between 1998 and 2014
```

```r
hilomap(b, data.frame(b)[,c("p_pt1998","p_pt2002","p_pt2006","p_pt2010","p_pt2014")],
        border="transparent", frame = e)

# PT using one standard deviation as lower limit to exclude those cases too close
# to the mean
par(mar=rep(0,4))
hilomap(b, data.frame(b)[,c("p_pt2006","p_pt2010","p_pt2014")], border="transparent",
        frame = e, zlim=1)
```




```r
par(mfrow=c(1,1))
```

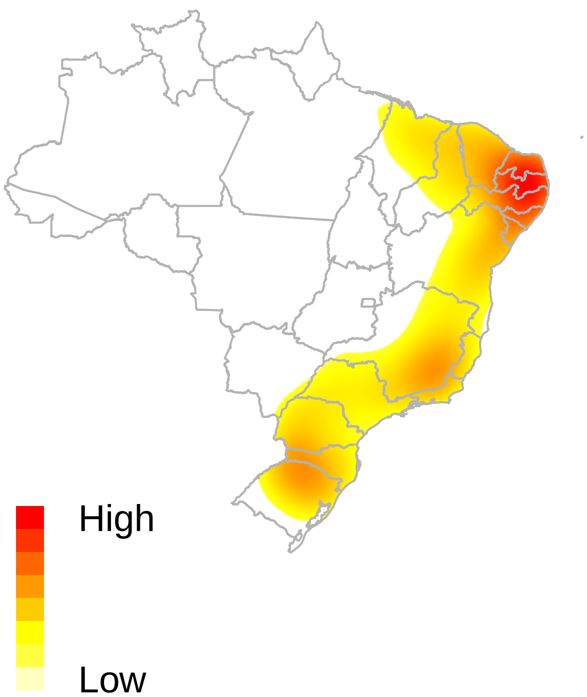
Kernel Density Map or Hot-Spot Map

Another common mapping application is to identify hot-spots or areas of particular high levels of concentration of events. These maps consider the density of the ocurrence of a given phenomenon and represents the results in a continuous surface where colors range from higher to lower density values.

In this example we map the density of percentage of votes in the PT and the PSDB in 2014 presidential elections. As we can see, the PT is concentrated in the Atlantic Coast with some more pronounced clusters in the Northeast, the east of Minas Gerais, and in the southern border, next to Argentina. The PSDB concentrates its votes mostly in the center-South with little presence in the Northeast coast.

```r
b$qt_pt2014[is.na(b$qt_pt2014)] <- 0
b$qt_psdb2014[is.na(b$qt_psdb2014)] <- 0
b$p_pt2014[is.na(b$p_pt2014)] <- 0
b$p_psdb2014[is.na(b$p_psdb2014)] <- 0
```

```r
# Density of the percentage of votes in PT and PSDB
pt <- dotsInPolys(pl = b, x = b$p_pt2014*100)
ps <- dotsInPolys(pl = b, x = b$p_psdb2014*100)

par(mfrow=c(1,2))
par(mar=rep(0,4))
kernelmap(pt, e, e, alpha=1.5)
kernelmap(ps, e, e, alpha=1.5, draw.leg = F)
```



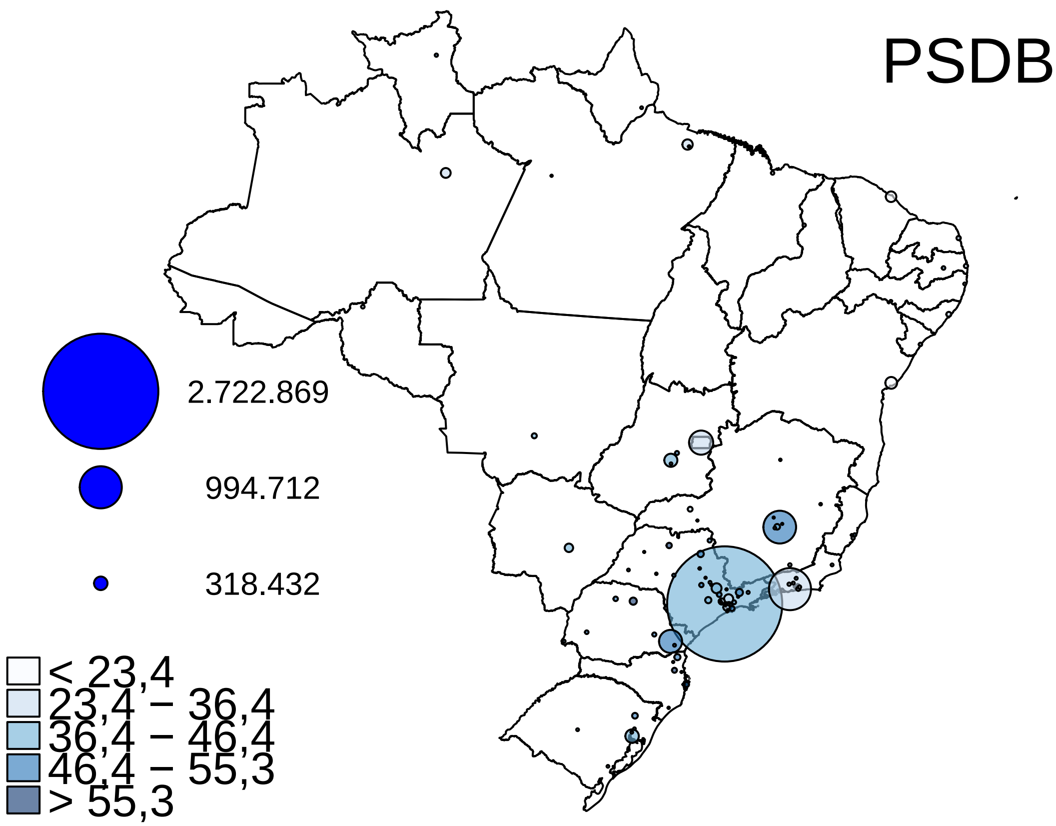

```r
par(mfrow=c(1,1))
```

Proportional Symbol Map

The proportional symbol is a widely used map. There is no need for further description. Our contribution here was to introduce some functionalities that allow the map to represent two variables at the same time. The first one will be symbolized by different circle sizes, while the second with a sequential color scheme. In our example, we will employ both absolute vote and percentage of vote in the PT in the 2014 elections.

```r
par(mar=rep(0,4))
propmap(b[b$qt_psdb2014>50000,], b$qt_psdb2014[b$qt_psdb2014>50000], frame = e,
        varclass = b$p_psdb2014[b$qt_psdb2014>50000]*100, leg.aux = "bottomleft",
        leg.pos=c(-78.5,-10), max.size = 8, y.int = 2.5, x.int = 3, leg.cex = 1,
        leg.aux.cex = 1.4, symbol.bg = "blue",leg.brks = 2, transp=98, y.int.aux = 0.6)
```

```
text(x = -39,y = 2.9, "PSDB", adj = 0, cex = 2)
```



Flow Map

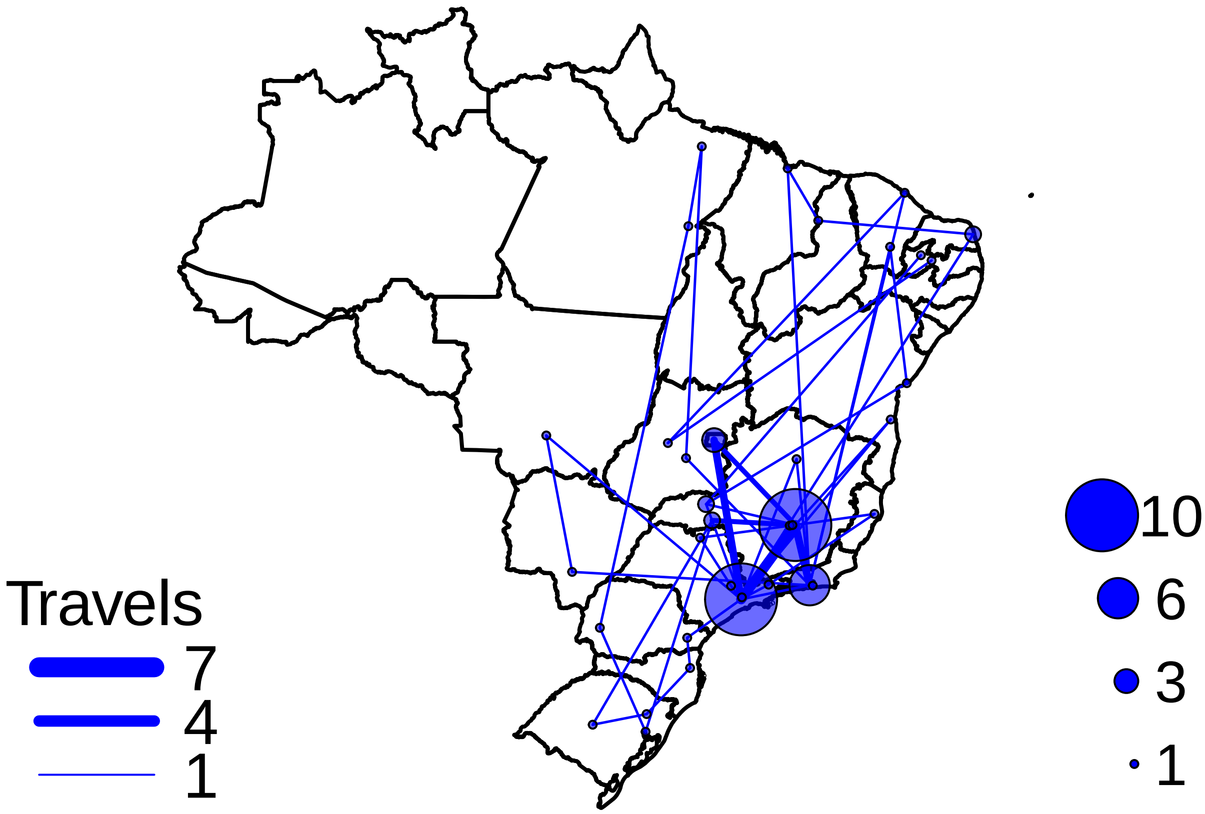
Another important representation is related to flows. How many times a candidate has gone from a city to another? What are the trajectories made by candidates in the last two or three months of the campaign? For answering these questions, we developed a set of functions that associate areas (regiones, cities, localities) to a latitude and longitude using their codes and, then, to create a spatial Lines object containing the trajectories. Using data from newspapers, we followed the two main contenders - Dilma Rousseff and Aécio Neves - in the last three months of their campaigns. The results are shown in the maps below:

```
cid <- read.csv("ciudades.csv", header=T, sep=";")
cd <- cid
cid  <- cid[,c("GEOCODIG_M","nome")]

mov <- read.csv("Mov_2014.csv", header=T, sep=";")

dtb <- merge(mov, cid, by.x="origin", by.y="nome", all.x=T, sort=F)
names(dtb)[5] <- "COD_OR"

dtb <- merge(dtb, cid, by.x="destination", by.y="nome", all.x=T, sort=F)
names(dtb)[6] <- "COD_DEST"


dtd <- dtb[dtb$candidate=="Dilma",]
```

```r
dta <- dtb[dtb$candidate=="Aécio",]


mDi <- calcMigra(dtd, "COD_OR", "COD_DEST", "total")
mAe <- calcMigra(dta, "COD_OR", "COD_DEST", "total")




# Finds the coordinates for the Brazilian cities
bx <- br[br$year==2014 & br$party==13,]

# Create the spatial flows object for each candidate
pDi <- ptCoords(bx,idvar = "GEOCODIG_M", origin = mDi$origin,
                destination = mDi$destination, spline = T, data = mDi)
pAe <- ptCoords(bx,idvar = "GEOCODIG_M", origin = mAe$origin,
                destination = mAe$destination, spline = T, data = mAe)


pDi$lwd <- pDi$total_flow/max(pDi$total_flow)*5
pAe$lwd <- pAe$total_flow/max(pAe$total_flow)*5

bDi <- merge(bx, cd[,c("GEOCODIG_M","nome","dilma")])
bAe <- merge(bx, cd[,c("GEOCODIG_M","nome","aecio")])

bDi <- bDi[which(bDi$dilma>0),]
bDi <- SpatialPointsDataFrame(bDi, data = data.frame(bDi))
bDi$lwd <- round((bDi$dilma/max(bDi$dilma))*10,1)

bAe <- bAe[which(bAe$aecio>0),]
bAe <- SpatialPointsDataFrame(bAe, data = data.frame(bAe))
bAe$lwd <- round((bAe$aecio/max(bAe$aecio))*10,1)
```

```r
par(mar=rep(0,4))
plot(e, lwd=2)
plot(pDi, add=T, lwd=pDi$lwd, col="red")
propmap(bDi, bDi$lwd, add=T, col="red", trans=95, leg.pos="bottomright", y.int=1.2)
legend("bottomleft", col="red", lwd=c(10,5.7,1),legend=c(7,4,1), bty="n",
       title="Travels", cex=2, y.intersp=0.7, x.intersp=0.5,title.adj=0.35 )
```


```r
par(mar=rep(0,4))
plot(e, lwd=2)
plot(pAe, add=T, lwd=pAe$lwd, col="blue")
propmap(bAe, bAe$lwd, add=T, col="blue", symbol.bg = "blue",trans=95,
        leg.pos="bottomright", y.int=1.2)
legend("bottomleft", col="blue", lwd=c(10,5.7,1),legend=c(7,4,1), bty="n",
       title="Travels", cex=2, y.intersp=0.7, x.intersp=0.5,title.adj=0.35 )
```


Travels
7
4
1
10
6
3
1