



# MÁSTER DE ESTADÍSTICA APLICADA CON R SOFTWARE

Descripción, correlación y análisis espacial de la  
población de estudiantes salvadoreños reprobados de  
2015 a nivel municipal

AUTOR: Felipe Antonio Carranza Flores

DIRECTOR: Juan Luis López Garrancho

FECHA: 13 de abril 2020

ENTIDAD COLABORADORA:





# Resumen

El porcentaje de alumnos reprobados en la educación básica para el 2015 en El Salvador a nivel municipal se comporta diferente a nivel descriptivo en las categorías de sexo y turno, existe evidencia de grupos autocorrelacionados espacialmente significativamente estadísticos principalmente en la zona occidental y norte del país. El porcentaje total está en su mayoría relacionado con el turno diurno y el sexo masculino, hay municipios con altos o bajos porcentajes no influidos por sus vecinos, el PIB per cápita municipal calculado se descartó como variable explicativa lineal mediante regresión. Debido a la falta de informes de alumnos reprobados desde el 2014 a nivel municipal y la ausencia de estadística geográfica a nivel histórico en la información de los portales públicos, se procesó con software R la base de datos del Ministerio de Educación y tecnologías, así como los datos económicos y geográficos obtenidos por investigadores sociales en el país. Se hizo una descripción de los porcentajes de reprobados a nivel total, por sexo y turnos, se encontraron correlaciones no geográficas, se generaron mapas descriptivos y pruebas de autocorrelación global con el test de Índice de Morán y mapas de clusters a partir de los test de Getis Ord.

*Palabras clave:* educación básica, reprobados, El Salvador, autocorrelación.

# Agradecimientos

A mis tutores del Máster en R Juan Luis López y Rosario Ferrero, al personal de Máxima Formación, al economista César Sánchez por compartirme sus datos propios del PIB per cápita municipal del El Salvador, a la especialista en Sistemas de Información Geográfica (SIG) Metzi Aguilar por facilitarme los datos de ubicación geográfica de los municipios con puntos y límites geográficos que pone a disposición el Centro Nacional de Registros (CNR), a mi familia por el apoyo incondicional.

# Índice

<b>Índice de figuras</b>	<b>4</b>
<b>Índice de Tablas</b>	<b>5</b>
<b>Introducción</b>	<b>6</b>
<b>Material y métodos</b>	<b>7</b>
Aportaciones teóricas	7
Técnicas cuantitativas	7
<b>Resultados</b>	<b>8</b>
<b>Discusión</b>	<b>28</b>
<b>Conclusiones</b>	<b>34</b>
<b>Referencias</b>	<b>36</b>
<b>Anexo A: código fuente en R</b>	<b>37</b>

## Índice de figuras

- [Figura 1. Correlación entre porcentaje de Total de reprobados y el porcentaje del Total de reprobados sex masc. cor= 0.9632285](#)
- [Figura 2. Correlación entre el porcentaje de Total de reprobados y el porcentaje del Total de reprobados sex fem. cor= 0.879185](#)
- [Figura 3. Correlación entre el porcentaje de Total de reprobados y el Total de reprobados turno diurno . cor= 0.9975032](#)
- [Figura 4. Correlación entre porcentaje de Total de reprobados y el el porcentaje del Total de reprobados turno diurno .](#)
- [Figura 5. Correlación entre el porcentaje de Total de reprobados y el porcentaje del Total de reprobados turno distancia .](#)
- [Figura 6. Mapa del porcentaje del total de reprobados a nivel municipal](#)
- [Figura 7. Mapa del porcentaje del total de reprobados del sexo masculino a nivel municipal](#)
- [Figura 8. Mapa del porcentaje del total de reprobados del sexo femenino a nivel municipal.](#)
- [Figura 9. Mapa del porcentaje del total de reprobados del turno diurno a nivel municipal.](#)
- [Figura 10. Mapa del porcentaje del total de reprobados del turno nocturno a nivel municipal](#)
- [Figura 11. Mapa del porcentaje del total de reprobados del turno distancia a nivel municipal.](#)
- [Figura 12. Mapa del listado de vecindarios resultantes de la matriz de distancias entre municipios \( `knn2nb\(knearneigh\(coords, k = 4\) \)` \)](#)
- [Figura 13. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de reprobados a nivel municipal, 95% de confianza.](#)
- [Figura 14. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de reprobados de sexo masculino a nivel municipal, 95% de confianza.](#)
- [Figura 15. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de reprobados de sexo femenino a nivel municipal, 95% de confianza](#)
- [Figura 16. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de reprobados de turno diurno a nivel municipal, 95% de confianza.](#)
- [Figura 17. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de reprobados de turno nocturno a nivel municipal.](#)
- [Figura 18. Clustering alto/bajo \(G general de Getis-Ord\) porcentaje del total de](#)

- [reprobados de turno distancia a nivel municipal, 95% de confianza.](#)
- [Figura 19. Porcentaje del total de reprobados.](#)

## Índice de Tablas

- [Tabla 1. Sumario de la variable total de reprobados por municipio para el año 2015](#)
- [Tabla 2. Sumario de la variable total de reprobados por municipio del sexo masculino para el año 2015](#)
- [Tabla 3. Sumario de la variable total de reprobados por municipio del sexo femenino para el año 2015](#)
- [Tabla 4. Sumario de la variable total de reprobados por municipio del turno diurno para el año 2015](#)
- [Tabla 5. Sumario de la variable total de reprobados por municipio del turno nocturno para el año 2015](#)
- [Tabla 6. Sumario de la variable total de reprobados por municipio del turno distancia para el año 2015](#)
- [Tabla 7. Razón de sexo \(masculino / femenino\) de los estadísticos: mediana y media](#)
- [Tabla 8. Razón de turno \(diurno / nocturno \) de la media, el turno nocturno tiene mediana = 0](#)
- [Tabla 9. Razón de turno \(diurno / distancia \) de la media, el turno distancia tiene mediana = 0](#)
- [Tabla 10. Razón de turno \(nocturno / distancia \) de la media](#)

# Introducción

En El Salvador, no se dispone publicaciones a nivel municipal del porcentaje de alumnos reprobados en educación básica por lo que se carece de un registro de indicadores en educación en relación al rendimiento escolar que permita realizar investigaciones sobre la calidad de la educación en el país. La información oficial gubernamental de los portales de internet del Ministerio de Educación de El Salvador no tiene disponibles publicaciones a nivel municipal donde se visualice y analice dicha información desde el año 2014 (Anuario Estadístico Educativo Centros Educativos Regulares y de Educación Especial 2014). Aunque estos datos de tipo cuantitativo de la población de reprobados por municipio existen en las bases de datos de censo de matrícula final de dicho ministerio desde 2005 al 2018, en las publicaciones solo se hace referencia a estos por categorías de agrupación más grande que la de municipio como son los departamentos. Tampoco se dispone de visualizaciones de esta variable en mapas y ni de las relaciones geográficas entre los municipios con la variable de interés. Además, no existen estudios que busquen explicar relaciones entre el nivel de reprobados y las variables económicas.

El objetivo de este estudio es procesar el censo de matrícula final 2015 y describir a nivel cuantitativo la población de reprobados por municipio, turnos y sexo y las posibles correlaciones entre las variables. Visualizar dichos porcentajes en mapas para tener una visión geográfica del fenómeno así como encontrar posibles relaciones geográficas entre los niveles de reprobados y la ubicación espacial entre los municipios. Encontrar si hay relaciones entre nivel de reprobados y una variable proxy económica asociada.

# Material y métodos

## Aportaciones teóricas

### Material:

1. [Base de datos Datos de matrícula final, censo 2015 del Ministerio de salud](#)
2. Valores propios calculados de PIB per cápita 2015 por municipio (Economista César Sánchez)
3. Límites y puntos centrales geográficos de los municipios (Especialista SIG Metz Aguilar). El Salvador cuenta con 262 municipios agrupados en 14 departamentos.

## Técnicas cuantitativas

**Descriptivo.** Para determinar las estadísticas descriptivas de los porcentajes de reprobados por municipio y por las categorías turno y sexo,

**Correlacional** Para determinar la correlación entre el porcentaje de total de reprobados y las categorías turno y sexo.

**Descriptivo espacial.** Para visualizar en mapas las estadísticas descriptivas de los porcentajes de reprobados por municipio y por las categorías turno y sexo.

**Autocorrelación espacial** Para determinar si hay autocorrelación espacial en la variable de porcentaje del total de reprobados, las categorías de turno y sexo desde el enfoque geográfico.

**Clustering alto y bajo:** Para determinar si existen agrupamientos geográficos en la variable porcentaje del total de reprobados y las categorías de turno y sexo.

**Relación explicativa:** Para determinar si se puede explicar el porcentaje del total reprobados con el PIB per cápita nivel municipal.

## Resultados

### Resultado 1 Datos descriptivos, Total de registros del censo:

A. Resumen de los porcentajes del total de reprobados por municipio:

*Tabla 1. Sumario de la variable total de reprobados por municipio para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	3.070%	3.380%	13.514%	1.460553

B. Resumen de los porcentajes del total de reprobados por municipio del sexo masculino:

*Tabla 2. Sumario de la variable total de reprobados por municipio del sexo masculino para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	1.965%	2.229%	11.486%	0.9237447

C. Resumen de los porcentajes del total de reprobados por municipio del sexo femenino:

*Tabla 3. Sumario de la variable total de reprobados por municipio del sexo femenino para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	1.0526%	1.1605%	4.9755%	0.5995644

D. Resumen de los porcentajes del total de reprobados por municipio del turno diurno:

*Tabla 4. Sumario de la variable total de reprobados por municipio del turno diurno para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	3.007%	3.336%	13.514%	1.438573

E. Resumen de los porcentajes del total de reprobados por municipio del turno nocturno:

*Tabla 5. Sumario de la variable total de reprobados por municipio del turno nocturno para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	0.00%	0.02105%	0.63718%	0.0834030

F. Resumen de los porcentajes del total de reprobados por municipio del turno distancia:

*Tabla 6. Sumario de la variable total de reprobados por municipio del turno distancia para el año 2015*

Valor mínimo	Mediana	Media	Valor máximo	Desviación Estándar
0.00%	0.00%	0.02341%	1.19863 %	0.2047539

**Razones entre las categorías:**

- Sexo (masculino / femenino)

Tabla 7. Razón de sexo (masculino / femenino) de los estadísticos: mediana y media

Mediana	Media
1.866806	1.92

- Turno (Diurno / Nocturno)

Tabla 8. Razón de turno (diurno / nocturno) de la media, el turno nocturno tiene mediana = 0

Mediana	Media
N/A, Mediana Nocturno = 0	158.4798

- Turno (Diurno / Distancia)

Tabla 9. Razón de turno (diurno / distancia) de la media, el turno distancia tiene mediana = 0

Mediana	Media
N/A, Mediana distancia = 0	142.5032

- Turno (Nocturno / Distancia)

Tabla 10. Razón de turno (nocturno / distancia) de la media

Media
0.8991884

**Resultado 2 Correlaciones** para determinar la correlación entre el porcentaje de total de reprobados y las categorías turno y sexo.

A. Correlación entre Total de reprobados y el Total de reprobados de reprobados del sexo femenino

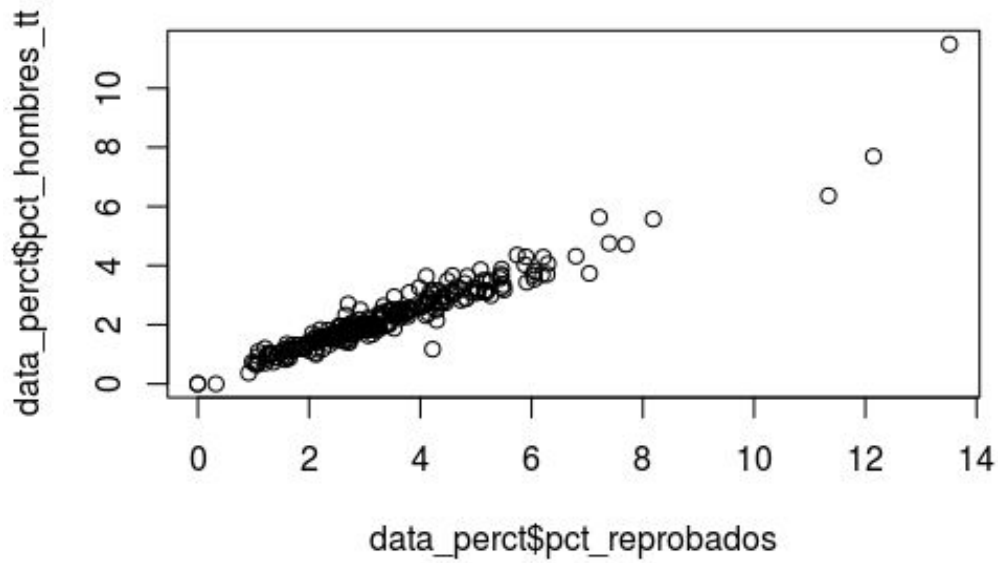


Figura 1. Correlación entre porcentaje de Total de reprobados y el porcentaje del Total de reprobados sex masc.  $cor= 0.9632285$

B. Correlación entre Total de reprobados y el Total de reprobados de reprobados del sexo femenino

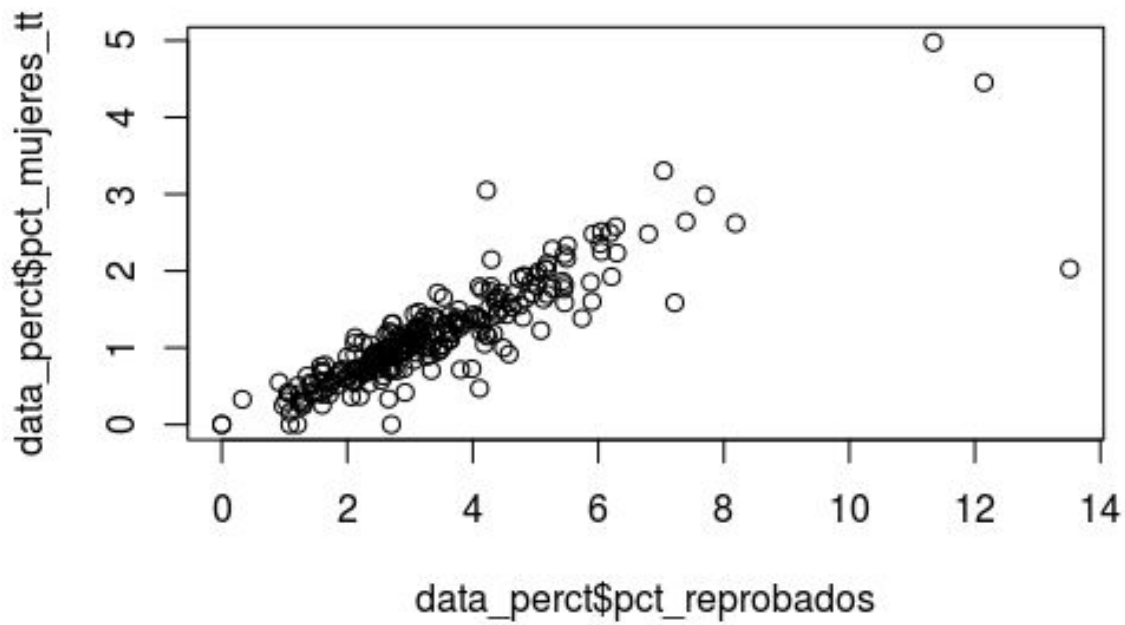


Figura 2. Correlación entre el porcentaje de Total de reprobados y el porcentaje del Total de reprobados sex fem.  $cor= 0.879185$

C. Correlación entre Total de reprobados y el Total de reprobados de reprobados del turno diurno

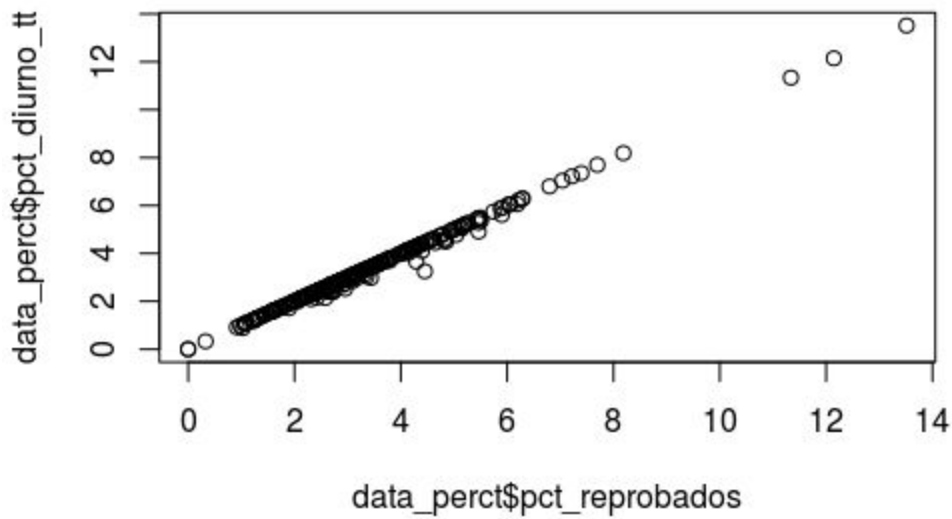


Figura 3. Correlación entre el porcentaje de Total de reprobados y el Total de reprobados turno diurno .  $cor= 0.9975032$

D. Correlación entre Total de reprobados y el Total de reprobados del turno nocturno

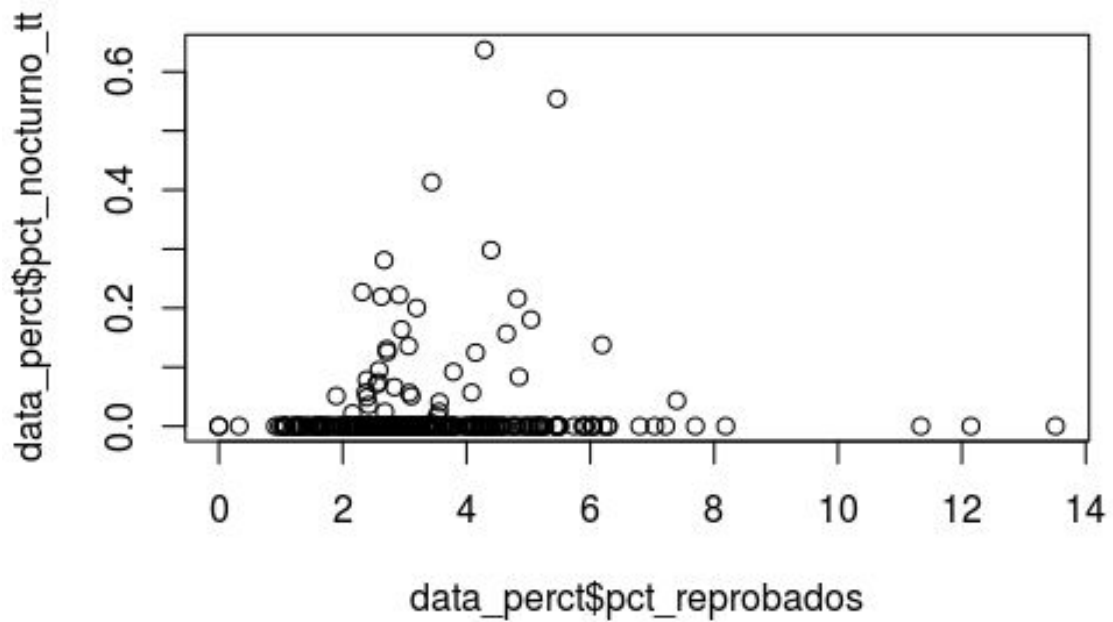


Figura 4. Correlación entre porcentaje de Total de reprobados y el el porcentaje del Total de reprobados turno diurno .

E. Correlación entre porcentaje de Total de reprobados y el Total de reprobados del turno distancia

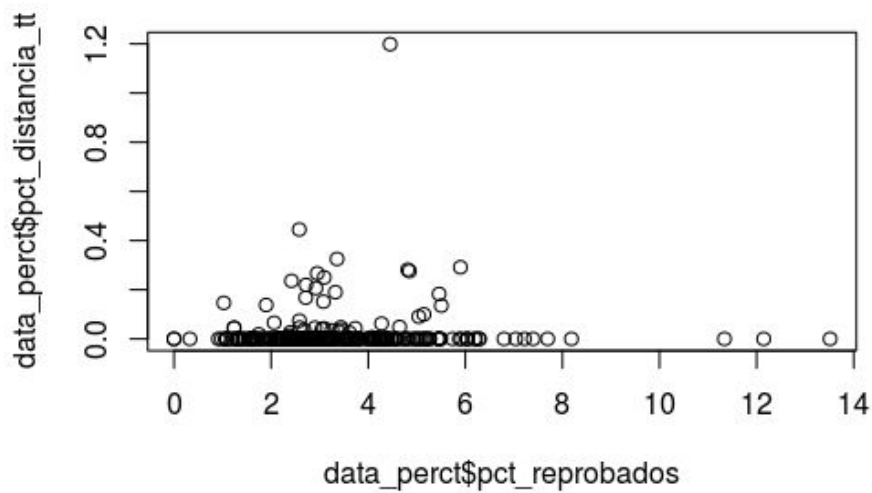


Figura 5. Correlación entre el porcentaje de Total de reprobados y el porcentaje del Total de reprobados turno distancia .

### Resultado 3 Descriptivo espacial.

Para visualizar en mapas los estadísticas descriptivas de los porcentajes de reprobados por municipio y por las categorías turno y sexo.

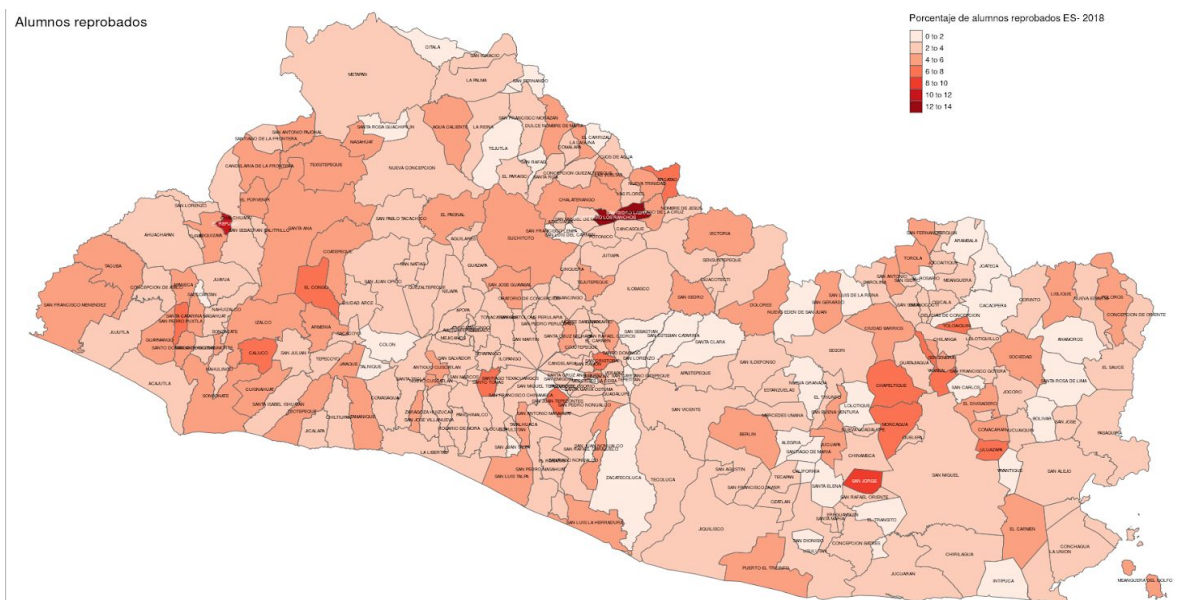


Figura 6. Mapa del porcentaje del total de reprobados a nivel municipal

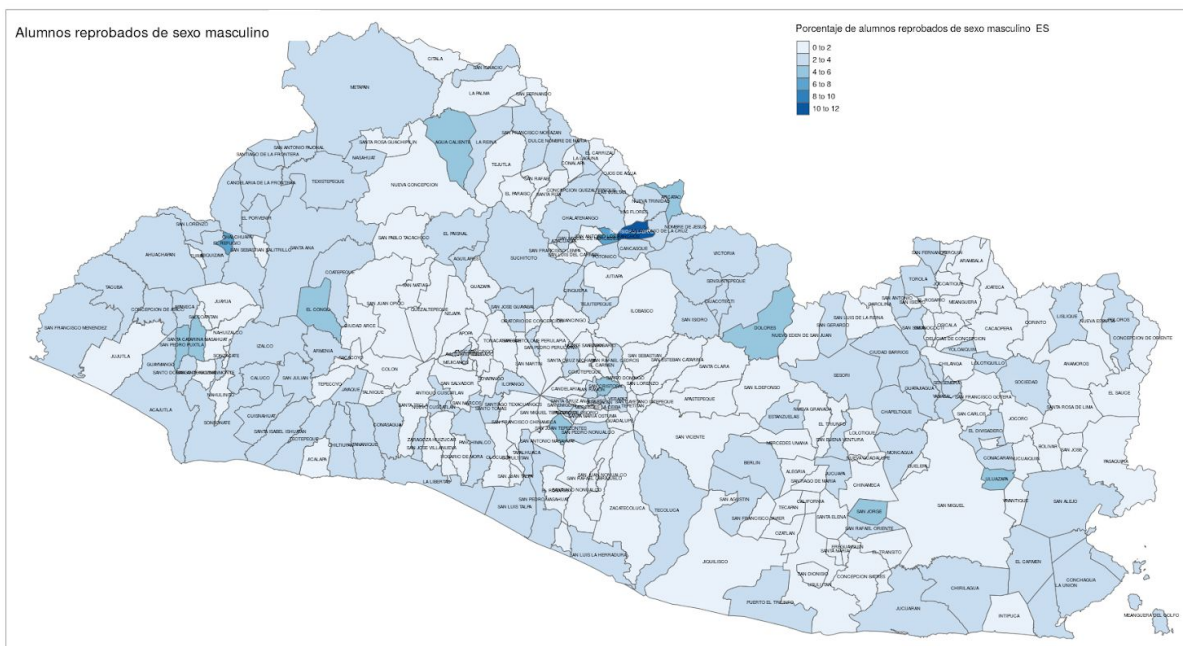


Figura 7. Mapa del porcentaje del total de reprobados del sexo masculino a nivel municipal

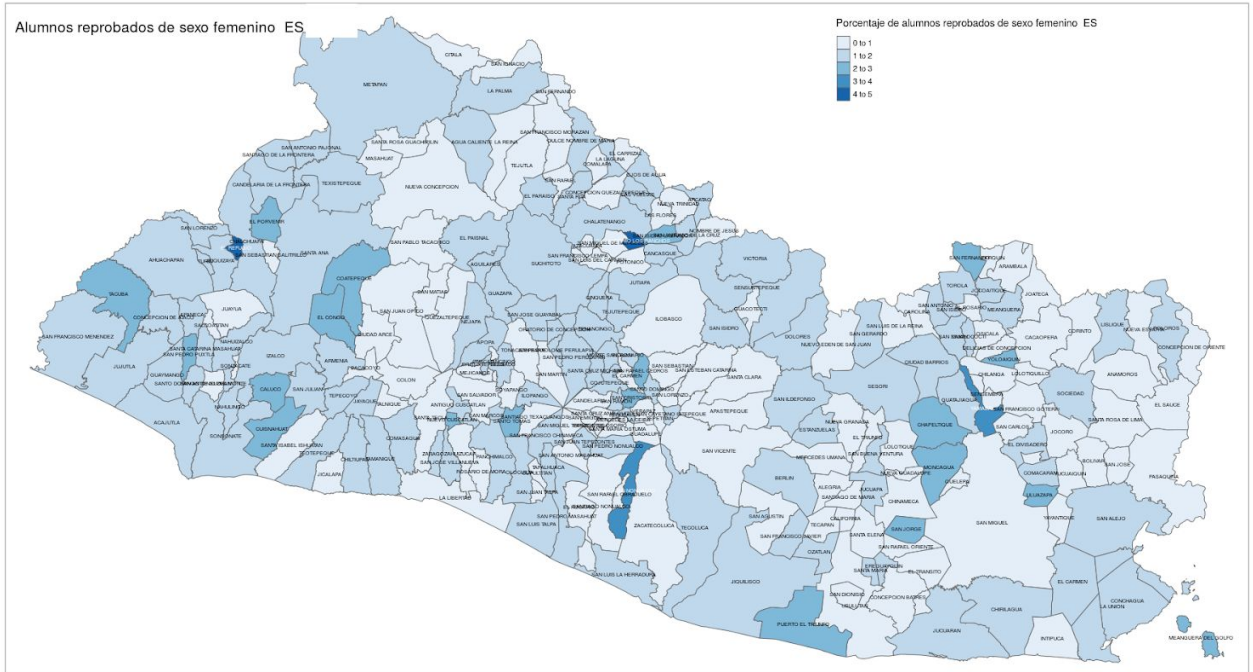


Figura 8. Mapa del porcentaje del total de reprobados del sexo femenino a nivel municipal.

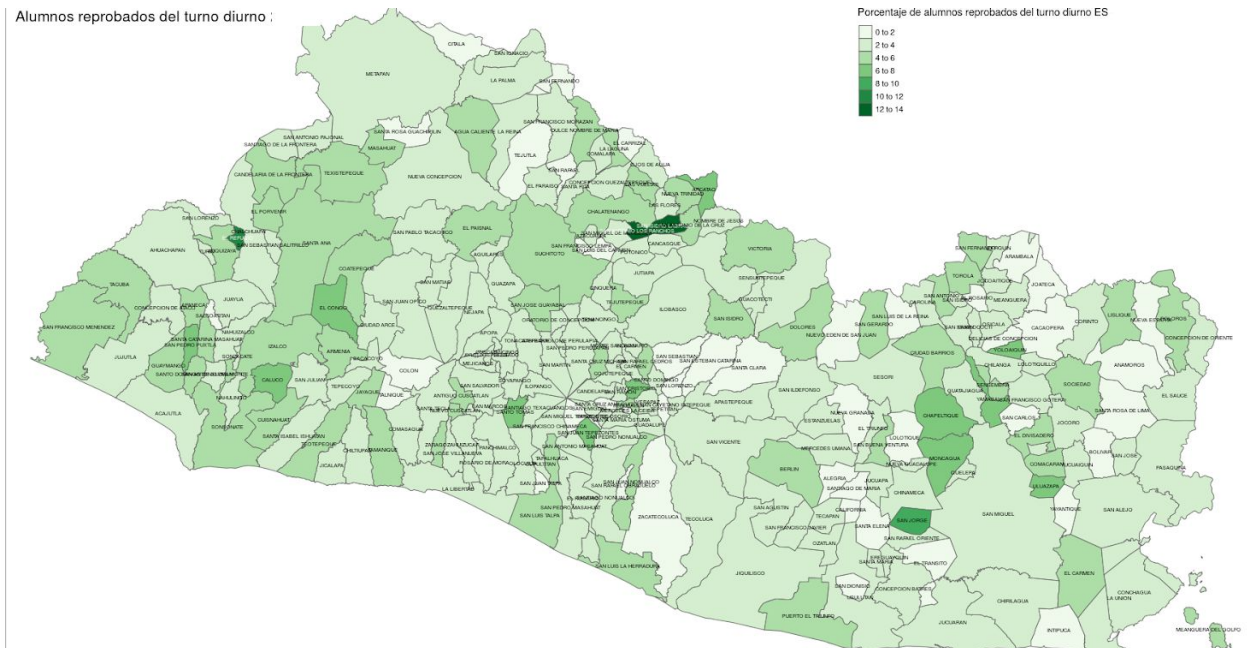


Figura 9. Mapa del porcentaje del total de reprobados del turno diurno a nivel municipal.

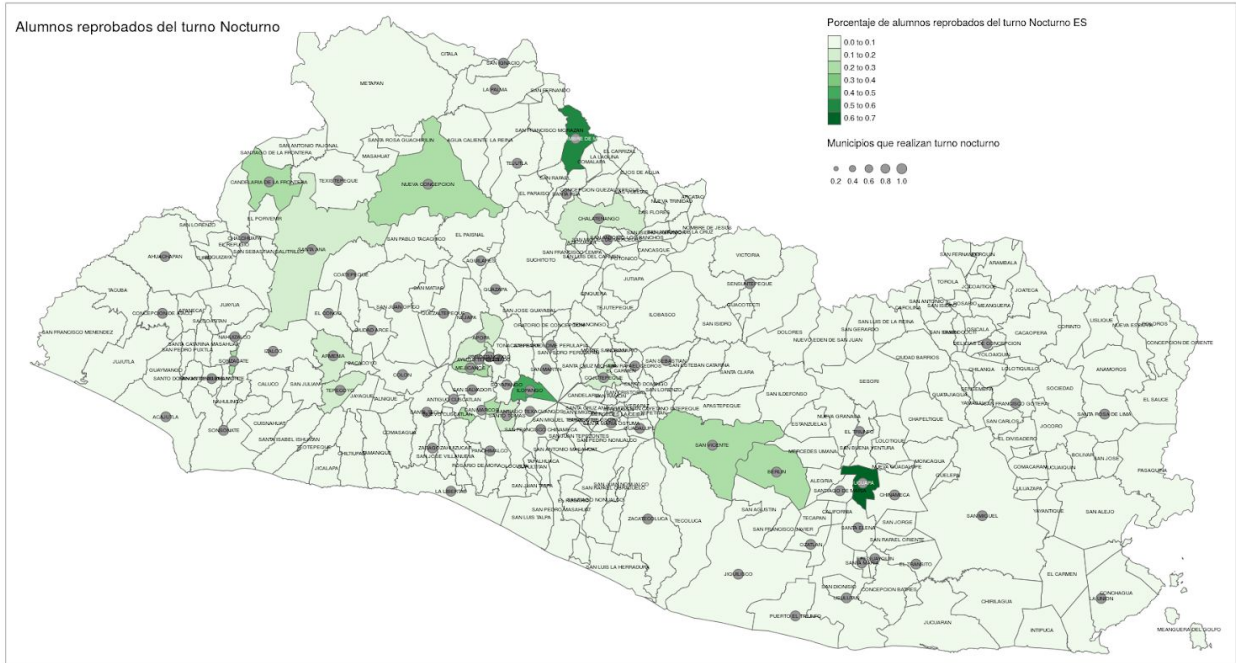


Figura 10. Mapa del porcentaje del total de reprobados del turno nocturno a nivel municipal



Figura 11. Mapa del porcentaje del total de reprobados del turno distancia a nivel municipal.

**Resultado 4 Autocorrelación espacial:** Para determinar si hay autocorrelación espacial en la variable de porcentaje del total de reprobados, las categorías de turno y sexo desde el enfoque geográfico.

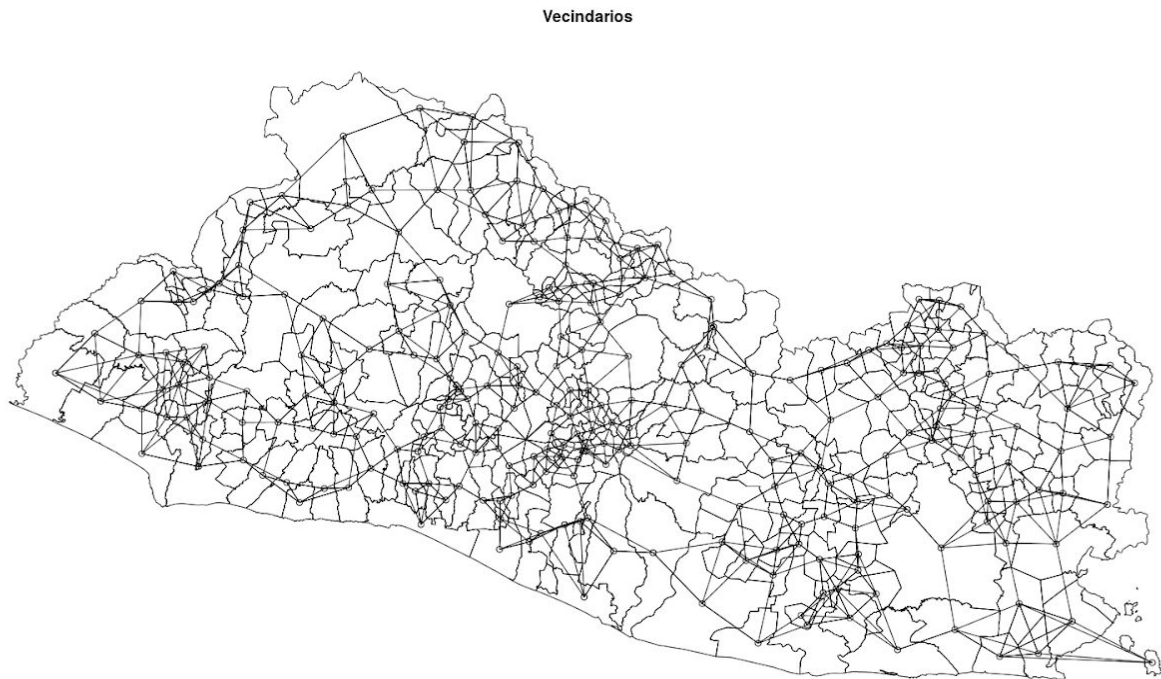


Figura 12. Mapa del listado de vecindarios resultantes de la matriz de distancias entre municipios (  $knn2nb(knearneigh(coords, k = 4))$  )

Prueba de I Morán global:

- A. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.009611

- B. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de sexo masculino (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de sexo masculino no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.001255

- C. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de sexo femenino (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados del sexo femenino no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.1694

- D. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados del turno diurno (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados del turno diurno es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.01138

- E. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados del turno nocturno (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados del turno nocturno es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.1181

- F. Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados del turno distancia (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados del turno distancia es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.7713

**Resultado 5 Clustering alto y bajo:** Para determinar si existen agrupamientos en geográficos en la variable porcentaje del total de reprobados y las categorías de turno y sexo.

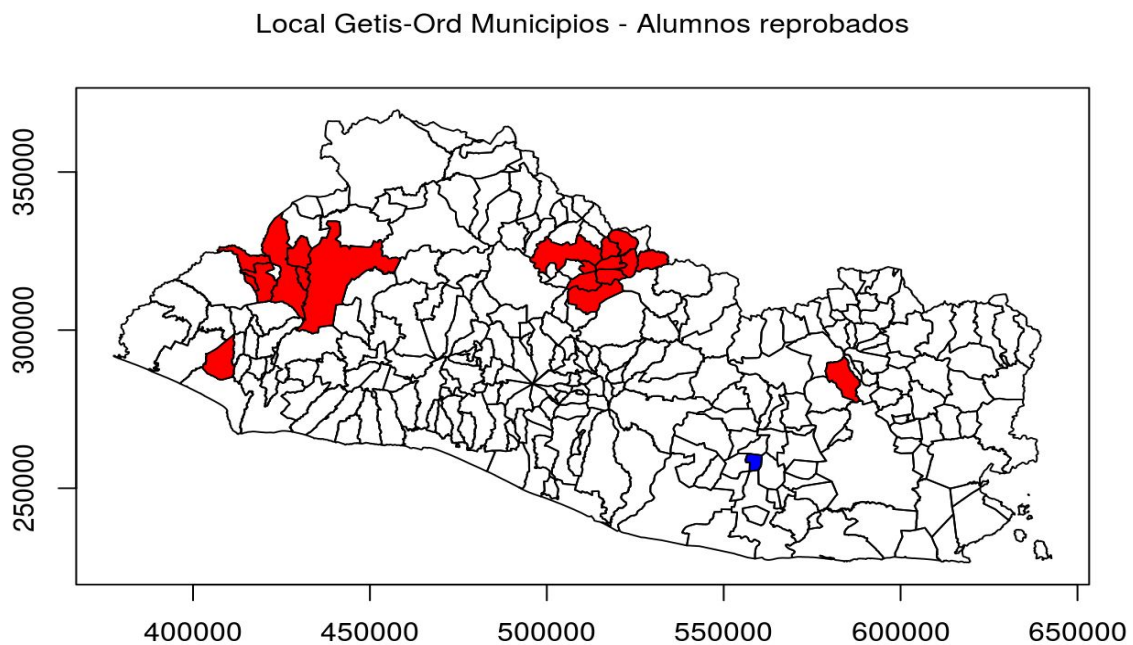


Figura 13. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados a nivel municipal, 95% de confianza.

Local Getis-Ord Reprobados Sexo masculino

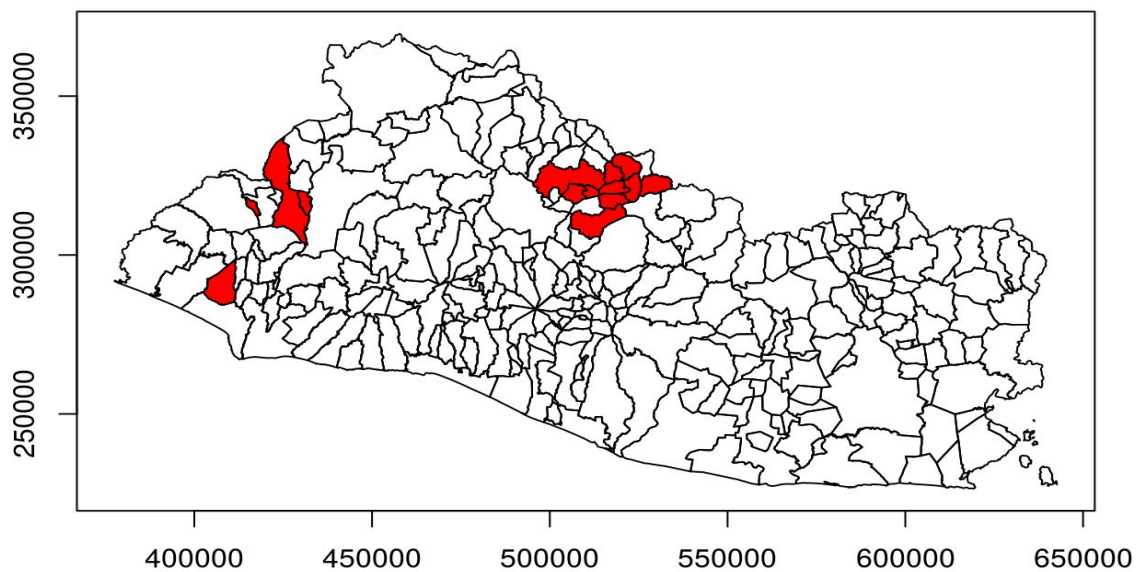


Figura 14. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de sexo masculino a nivel municipal, 95% de confianza.

Local Getis-Ord Reprobados Sexo Femenino

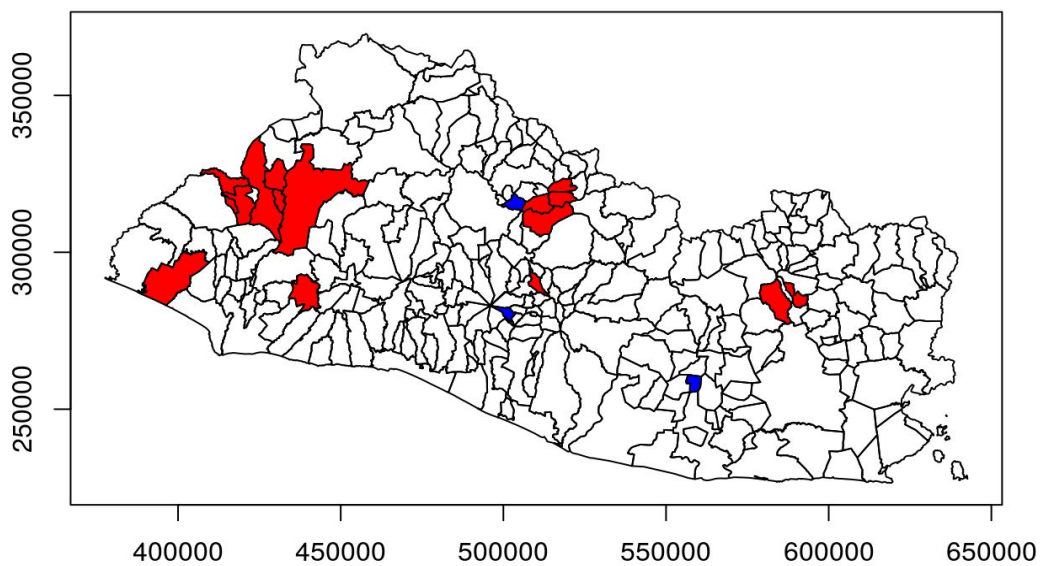


Figura 15. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de sexo femenino a nivel municipal, 95% de confianza

Local Getis-Ord Reprobados Turno diurno

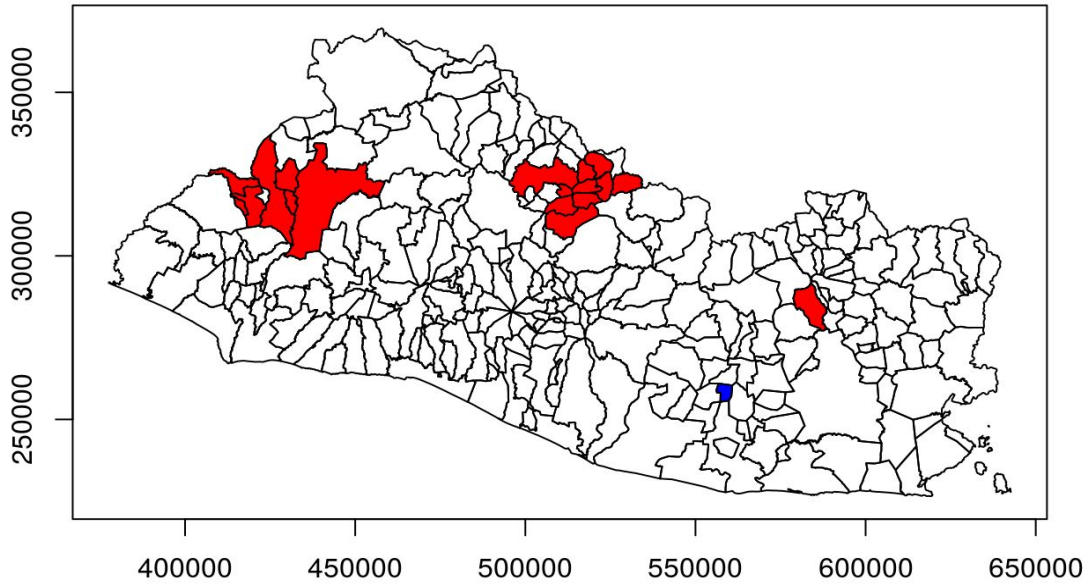


Figura 16. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno diurno a nivel municipal, 95% de confianza.

Local Getis-Ord Reprobados Turno nocturno

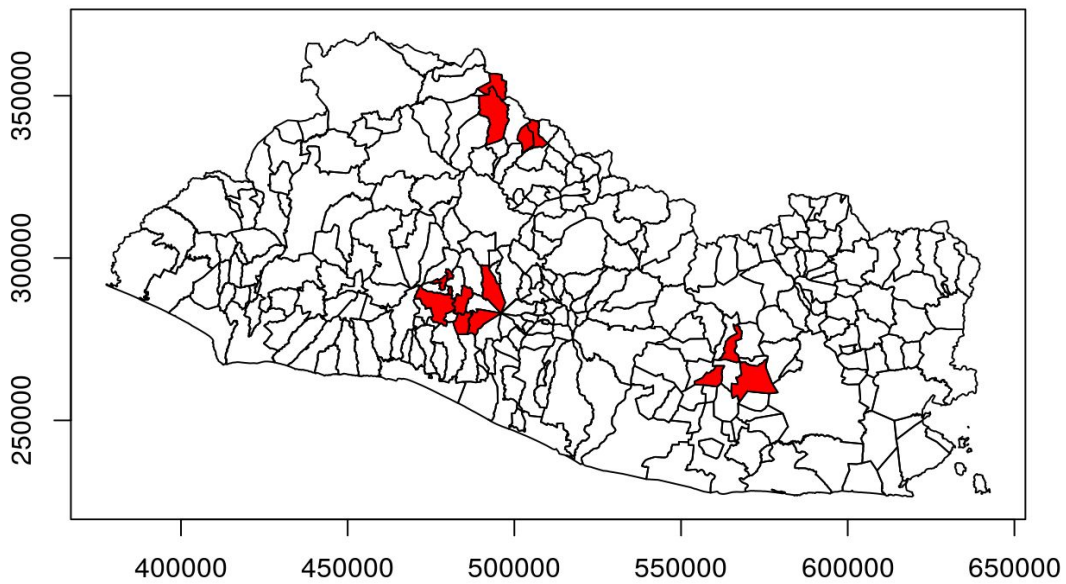


Figura 17. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno nocturno a nivel municipal.

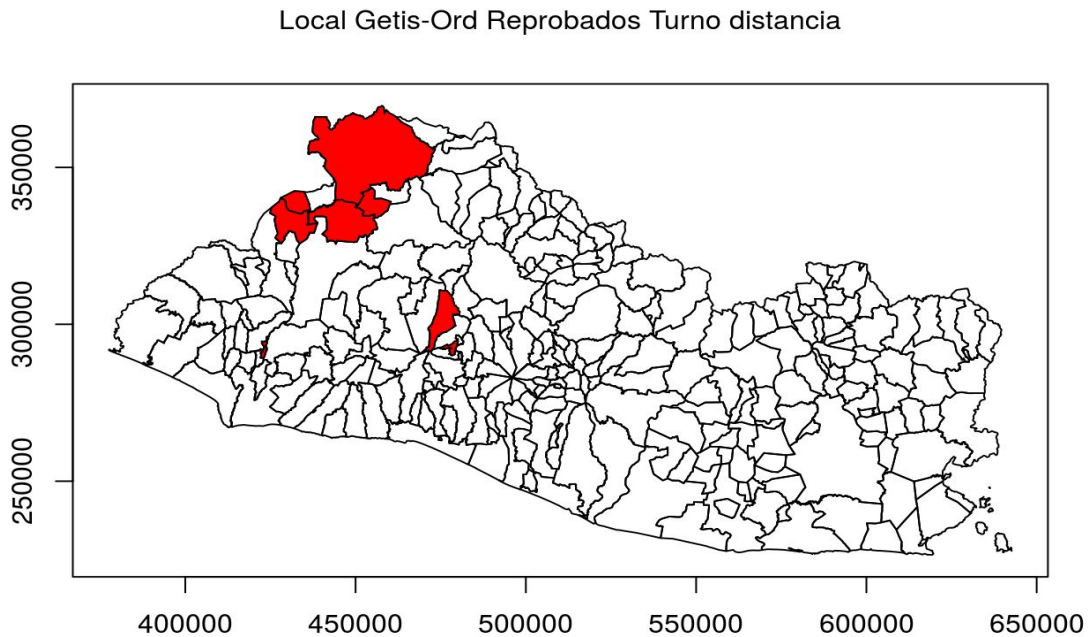


Figura 18. Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno distancia a nivel municipal, 95% de confianza.

**Resultado 6 Relación explicativa:** Para determinar si se puede explicar el porcentaje el reprobados en total y por categorías con el PIB per cápita nivel municipal.

- Variable PIB per cápita: PIB Per cápita para el 2015 por municipio
- Categorización de variable PIB per cápita con método de árbol de regresión
  - Categorías encontradas: 10
    - PIBPC\_2015 < 381.19 ~ 1,
    - PIBPC\_2015 >= 381.19 & PIBPC\_2015 < 692.7386 ~ 2,
    - PIBPC\_2015 >= 692.7386 & PIBPC\_2015 < 1060.156 ~ 3,
    - PIBPC\_2015 >= 1060.156 & PIBPC\_2015 < 1193.496 ~ 4,
    - PIBPC\_2015 >= 1193.496 & PIBPC\_2015 < 2481.467 ~ 5,
    - PIBPC\_2015 >= 2481.467 & PIBPC\_2015 < 4050.064 ~ 6,

- PIBPC\_2015 >= 4050.064 & PIBPC\_2015 < 5073.224 ~ 7,
- PIBPC\_2015 >= 5073.224 & PIBPC\_2015 < 6049.284 ~ 8,
- PIBPC\_2015 >= 6049.284 & PIBPC\_2015 < 6049.284 ~ 9,
- PIBPC\_2015 >= 6049.284 ~ 10

- Porcentaje del total de reprobados

- Procedimiento 1

- Variable independiente; PIB per cápita por municipios
- Variable dependiente: porcentaje del total de reprobados por municipio
- Método: Regresión lineal
- Resultado R2 ajustado: -0.003695
- Gráfico Resultante:

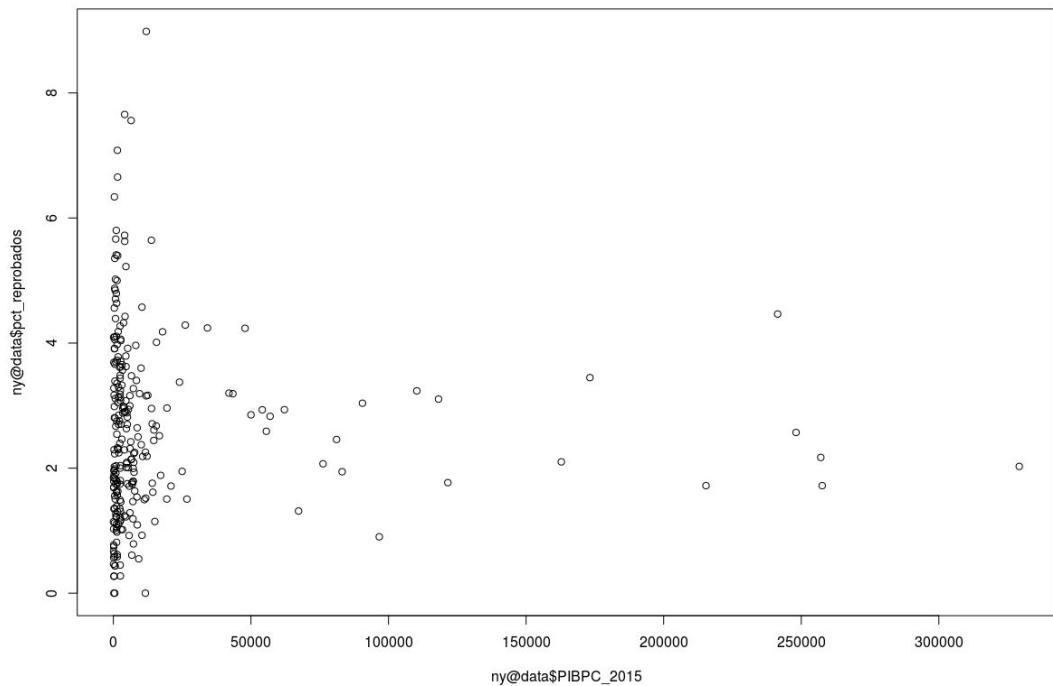


Figura 19. Porcentaje del total de reprobados explicado por el PIB per cápita

- Procedimiento 2
  - Variable independiente; PIB per cápita por municipios categorizado (10 categorías)
  - Variable dependiente: categorías total de reprobados por municipio
  - R2 ajustado 0.05973
- Procedimiento 3, Pruebas robustas y transformación de las variables
  - Versión Robusta R2 ajustado: 0.005893
  - Versión GAMR2 ajustado: 0.015
- Procedimiento 4, Regresión espacial entre las variables
  - Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados explicado por el PIB per cápita (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.06495

- Porcentaje del total de reprobados de sexo masculino
  - Variable independiente; PIB per cápita por municipios categorizado
  - Variable dependiente: porcentaje del total de reprobados de sexo masculino por municipio
  - Método: Regresión lineal
  - Resultado R2 ajustado: 0.05111
  - Resultado versión robusta R2 ajustado 0.1044
  - Regresión espacial entre las variables

- Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de sexo masculino explicado por el PIB per cápita(No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de sexo masculino por explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.06004

- Porcentaje del total de reprobados de sexo femenino
  - Variable independiente; PIB per cápita por municipios categorizado
  - Variable dependiente: porcentaje del total de reprobados de sexo masculino por municipio
  - Método: Regresión lineal
  - Resultado R2 ajustado: 0.06458
  - Resultado versión robusta R2 ajustado 0.07243
  - Regresión espacial entre las variables
    - Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de sexo femenino explicado por el PIB per cápita(No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de sexo femenino por explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.2624, con el 95% de confianza se mantiene la hipótesis nula.

- Porcentaje del total de reprobados de turno diurno
  - Variable independiente; PIB per cápita por municipios categorizado
  - Variable dependiente: porcentaje del total de reprobados de turno diurno por municipio
  - Método: Regresión lineal
  - Resultado R2 ajustado: 0.05961
  - Resultado versión robusta R2 ajustado 0.0648
  - Regresión espacial entre las variables
    - Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de turno diurno explicado por el PIB per cápita (No existe autocorrelación espacial).

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de turno diurno explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial).

Resultado del Test global de I Morán: p-value = 0.109

- Porcentaje del total de reprobados de turno nocturno
  - Variable independiente; PIB per cápita por municipios categorizado
  - Variable dependiente: porcentaje del total de reprobados de turno nocturno por municipio
  - Método: Regresión lineal

- Resultado R2 ajustado: 0.01199
- Resultado versión robusta R2 ajustado no calculado
- Regresión espacial entre las variables
  - Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de turno nocturno explicado por el PIB per cápita (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de turno nocturno explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.1269

- Porcentaje del total de reprobados de turno distancia
  - Variable independiente; PIB per cápita por municipios categorizado
  - Variable dependiente: porcentaje del total de reprobados de turno distancia por municipio
  - Método: Regresión lineal
  - Resultado R2 ajustado: 0.002344
  - Resultado versión robusta R2 ajustado no calculado
  - Regresión espacial entre las variables
    - Hipótesis nula: Existe asociación aleatoria entre los municipios para la variable de porcentaje total de reprobados de turno distancia explicado por el PIB per cápita (No existe autocorrelación espacial)

Hipótesis Alternativa: La asociación entre los municipios para la variable de porcentaje total de reprobados de turno distancia explicado por el PIB per cápita no es aleatoria (existe autocorrelación espacial)

Resultado del Test global de I Morán: p-value = 0.9313

## Discusión

### Interpretación de Resultado 1 Datos descriptivos:

El valor mínimo de porcentaje de reprobados por municipio es de 0 y el máximo es de 13.514%, teniendo una mediana 3.070 y una desviación estándar de 1.46. Para el caso de los porcentajes de reprobados por municipio pero categorizados por sexo se observa que la población de sexo masculino tiene valores superiores a los de sexo femenino en los estadísticos de media, mediana y valor máximo, obteniendo una razón Masculino / Femenino en media de 1.92 y 1.866806 para la mediana. Por otra parte, para la categoría de tipo turno es la del turno diurno la que presenta los valores muy superiores en los estadísticos de media (ver tablas 8 y 9).

### Interpretación de Resultado 2 Correlaciones :

En el caso de los porcentajes de reprobados en las categorías de sexo tanto para la parte masculina y femenina tienen alta correlación lineal con el porcentaje total, sin embargo en la parte de los turnos únicamente el turno diurno muestra alta correlación lineal con dicha variable.

### Interpretación de Resultado 3 Descriptivo espacial.

Los municipios que poseen los mayores valores de porcentajes de total de reprobados se encuentran en el área rural y son de la zona occidental, norte y oriental siendo estos: SAN ISIDRO LABRADOR, SAN ANTONIO DE LOS RANCHOS, EL REFUGIO, de igual manera los que tienen menor valor de reprobados se encuentran en el área rural, luego se puede observar una distribución no uniforme por departamento sino más bien agrupamientos geográficos donde la mayoría de municipios se observan entre el 2 y 4% de reprobados.

Para el caso de la categoría de sexo masculino coincide con los mismos municipios de porcentajes más altos que el porcentaje total de reprobados sin embargo la distribución de todo el país se encuentra sobre todo entre 0 y 2% de reprobados. Sin embargo, en los límites de la zona costera, norte y occidental los porcentajes varían sobre todo entre el 2 y 4% de reprobados.

Para el caso de la categoría de sexo femenino coincide con los mismos municipios de porcentajes más altos SAN ISIDRO LABRADOR, SAN ANTONIO DE LOS RANCHOS, pero aparece el YAMABAL en lugar del REFUGIO y bajos de reprobados que el porcentaje total de reprobados. Sin embargo, la distribución de todo el país se encuentra sobre todo entre 0 y 1% de reprobados, aunque en los límites de la zona costera, norte y occidental los porcentajes varían sobre todo entre el 1% y 2% de reprobados. Además, hay más presencias de reprobados entre 2 % y 3% en las diferentes zonas.

Para el caso de la categoría de turno diario los municipios que poseen los mayores valores de porcentajes de total de reprobados se encuentran en el área rural y son de

la zona occidental, norte y oriental coincidiendo con los del sexo masculino, por otra parte se observa un gran grupo de municipios de la zona occidental con valores de 4% a 8% mientras que la zona oriental excepto por 5 municipios con valores entre 6% y 8% la mayoría no pasan del 4%.

Para el caso de la categoría de turno nocturnos los municipios que poseen mayores porcentajes de total de reprobados son JUCUAPA, DULCE NOMBRE DE MARÍA e ILOPANGO, por otra parte la mayoría de municipios presenta rangos entre 0% y 0.1% de probados.

Para el caso de la categoría de turno distancia los municipios que poseen mayores porcentajes de total de reprobados son SAN ANTONIO PAJONAL, SAN ANTONIO MASAHUAT y SAN ISIDRO. Por otra parte, la mayoría de municipios presenta rangos entre 0% y 0.2% de reprobados.

Interpretación Resultado 4 Autocorrelación espacial: Para determinar si hay autocorrelación espacial en la variable de porcentaje del total de reprobados, las categorías de turno y sexo desde el enfoque geográfico.

Prueba de I Morán global:

Porcentaje total de reprobados:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.009611$ , con el 95% de confianza se rechaza la hipótesis nula (que la asociación aleatoria entre los municipios sea de tipo aleatoria), existiendo autocorrelación espacial.

Porcentaje total de reprobados de sexo masculino:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.001255$ , con el 95% de confianza **se rechaza la hipótesis nula** (que la asociación aleatoria entre los municipios sea de tipo aleatoria), existiendo autocorrelación espacial.

Porcentaje total de reprobados de sexo femenino:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.1694$ , con el 95% de confianza **se mantiene la hipótesis nula** (que la asociación aleatoria entre los municipios sea de tipo aleatoria).

Porcentaje total de reprobados del turno diurno:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.01138$ , con el 95% de confianza **se rechaza la hipótesis nula** (que la asociación aleatoria entre los municipios sea de tipo aleatoria), autocorrelación espacial.

Porcentaje total de reprobados del turno nocturno:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.1181$ , con el 95% de confianza **se mantiene la hipótesis nula** (que la asociación aleatoria entre los municipios sea de tipo aleatoria).

Porcentaje total de reprobados del turno distancia:

Con el valor de Test global de I Morán:  $p\text{-value} = 0.7713$ , con el 95% de confianza **se mantiene la hipótesis nula** (que la asociación aleatoria entre los municipios sea de tipo aleatoria).

Interpretación del Resultado 5 Clustering alto y bajo: Para determinar si existen agrupamientos geográficos en la variable porcentaje del total de reprobados y las categorías de turno y sexo.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados a nivel municipal al 95% de confianza:** Se observan dos clusters de valores altos, uno en la zona occidental y otra en la zona norte. Para el caso de la zona occidental se observa que EL REFUGIO a pesar de tener altos valores de reprobados estos no influyen ni son influidos por el cluster que lo rodea, diferente es en la zona norte donde el cluster contiene a SAN ISIDRO LABRADOR, SAN ANTONIO DE LOS RANCHOS, como valores atípicos se observan GUAYMANGO Y GUATAJIAGUA con valores altos y CALIFORNIA con valores bajos.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de sexo masculino a nivel municipal: al 95% de confianza:** Se observa un pequeño cluster de valores altos en la zona occidental entre CHALCHUAPA Y EL PORVENIR, el cluster de valores altos de la zona norte se asemeja bastante al del total de porcentajes de reprobados conteniendo a SAN ISIDRO LABRADOR y SAN ANTONIO DE LOS RANCHOS, se observan municipios aislados en la zona occidental pero ninguno de valores bajos.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de sexo femenino a nivel municipal al 95% de confianza:** Se observa un cluster de valores altos idéntico al de total de reprobados, para el caso del cluster de valores altos de la zona norte que incluye a SAN ISIDRO LABRADOR y SAN ANTONIO DE LOS RANCHOS se identifica el municipio de SAN LUIS DEL CARMEN como opuesto en valores bajos, el cluster de valores altos de la zona oriente está formado por GUATAJIAGUA y GUALOCOCTI.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno diurno a nivel municipal al 95% de confianza:** Se observa el mismo mapa que el de total de reprobados.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno nocturno a nivel municipal al 95% de confianza:** Todos los valores son de porcentajes altos, Se observan 2 cluster en la zona central el primero conformado por la capital del país SAN SALVADOR, SOYAPANGO, SANTO TOMÁS y SANTIAGO TEXACUANGO, el segundo por AYUTUXTEPEQUE y CUSCATANCINGO además de un municipio aislados: SAN MARTÍN, se observan dos cluster en el área norte cerca de la frontera con Honduras: SAN FRANCISCO MORAZÁN Y SAN FERNANDO, y el otro está conformado por LA LAGUNA Y COMALAPA, mientras que en la zona oriental se observa un cluster conformado por CHINAMECA y SAN BUENA VENTURA y un municipio aislado: SANTIAGO DE MARÍA.

**Clustering alto/bajo (G general de Getis-Ord) porcentaje del total de reprobados de turno distancia a nivel municipal al 95% de confianza:** Todos los valores son de porcentajes altos, Se observan 1 cluster en al zona occidental -norte conformado cerca de la frontera con Guatemala: METAPÁN, TEXISTEPEQUE, CANDELARIA LA FRONTERA, SANTIAGO DE LA FRONTERA, además se observan dos municipios aislados en la zona central: NEJAPA y AYUTUXTEPEQUE.

#### Interpretación de Resultado 6 Relación explicativa:

Para determinar si se puede explicar el porcentaje del total reprobados con el PIB per cápita nivel municipal.

**Para todas las variables (dependientes)** Porcentaje del total de reprobados, Porcentaje del total de reprobados de sexo masculino, Porcentaje del total de reprobados de sexo femenino.

Porcentaje del total de reprobados de turno diurno, Porcentaje del total de reprobados de turno nocturno, Porcentaje del total de reprobados de turno distancia se resume:

- Mediante la regresión lineal se obtuvo un R2 ajustado menor del 1%, pese al uso de transformaciones de las variables, pruebas robustas y/o

categorización de la variable independiente no se obtuvo mejoras por lo que se descarta una posible explicación de la variable dependiente

- Se verificó la relación espacial:
  - Resultado del Test global de I Morán:  $p\text{-value} > 0.05$ , **con el 95% de confianza se mantiene la hipótesis nula**; Existe asociación aleatoria entre los municipios para la variable dependiente explicado por el PIB per cápita (No existe autocorrelación espacial).

## Conclusiones

Los resultados e interpretaciones presentadas en este trabajo dan aportes para futuras investigaciones sobre el comportamiento de la población de reprobados a nivel escolar. En la fase descriptiva se logró observar que para cada categoría de sexo y turno existen diferentes comportamientos a nivel municipal abarcando en muchas de ellas zonas en el área rural. En general, la mayor parte de la población reprobada es representada y altamente correlacionada linealmente con el turno diurno y el sexo masculino, sin embargo es necesario evaluar a futuro las otras categorías y profundizar en la explicación del fenómeno. Así mismo, se detectaron grupos de municipios con autocorrelación espacial con valores altos y bajos diferentes según las categorías, la zona norte y occidental tienen municipios agrupados con altos valores de reprobados, para el caso del turno nocturno se puede apreciar la zona norte fronteriza y la zona central, mientras en el turno distancia la parte fronteriza occidental. También se detectaron municipios con altos porcentajes como el caso del REFUGIO que no está influenciado significativamente por el grupo que lo rodea, o el caso de en la categoría de sexo femenino que SAN LUIS DEL CARMEN es vecino de niveles bajos con un cluster de niveles altos, además se concluyó que no hay suficiente evidencia para

describir linealmente estas variables mediante la variable económica del PIB per cápita obteniendo bajísimos valores de  $R^2$  ajustado. De igual manera no se encontró evidencia que las regresiones hayan sido afectadas por una correlación espacial. Un problema grave en El Salvador es la falta de datos abiertos de tipo público, que permitan incluir nuevas variables relacionadas al fenómeno, y de no ser por investigadores que comparten sus datos proyectados o investigados sería aún más difícil realizar investigación social. Como parte de las mejoras a este estudio, se recomienda categorizar por año escolar. Finalmente, se concluye que el software R fue una herramienta indispensable para realizar este estudio al tener en un mismo software formas de procesar, evaluar y obtener información estadística de la data censal educativa, geográfica y económica.

## Referencias

Lilian Ramirez, (2015) Autocorrelación espacial: analogías y diferencias entre el índice de Moran y el índice Getis y Ord

Giuseppe Arbia, (2014) A primer for spatial econometrics with applications in R

Luc Anselin, (2005-2007) Spatial regression analysis in R a workbook

Jean-François Mas, (2018) Análisis espacial con R: Usa R como un Sistema de Información Geográfica

Máxima Formación, Tema 8. Estadística multivariada - Árbol de regresión multivariada (MRT).

Máxima Formación, Tema 4. Regresión y correlación.

Máxima Formación, Tema 2. Estadística descriptiva.

Ministerio de educación y tecnología de El Salvador, (2014) Anuario estadístico educativo centros educativos regulares y de educación especial.

## Anexo A: código fuente en R

```
#procesamiento del censo
datos = data.frame(read.csv("CF2015.csv"))

str(datos)

#Totales por variable
table(datos$CÓDIGO.DE.DEPARTAMENTO)
table(datos$MUNICIPIO)
table(datos$Zona)
table(datos$SECTOR)
table(datos$TURNO)
table(datos$GRADO)
#table(datos$SEXO)
table(datos$MUNICIPIO, )

#totales por municipio

library(dplyr)

library(tidyverse)
#install.packages("tidyverse")
library(tidyverse)

Por sexo turno

#str(datos)
#table(datos$REPROBADOS)
#str(datos$REPROBADOS.MASCULINOS)
#as.numeric(datos$REPROBADOS.MASCULINOS) +
as.numeric(datos$REPROBADOS.FEMENINOS)
datos = data.frame(read.csv("CF2015.csv"))
datos$REPROBADOS_TOTAL = (datos$REPROBADOS.MASCULINOS) +
(datos$REPROBADOS.FEMENINOS)
datos_temp2=NULL
datos_temp2 <- datos %>%
  group_by(MUNICIPIO, CÓDIGO.DE.MUNICIPIO) %>%
```

```

mutate(ttdistancia =sum(case_when(TURNO=="DISTANCIA" &
REPROBADOS_TOTAL> 0 ~ as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) ))
%>%
mutate(ttdiurno =sum(case_when(TURNO=="DIURNO" & REPROBADOS_TOTAL >
0 ~ as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) )) %>%
mutate(ttnocturno =sum(case_when(TURNO=="NOCTURNO" &
REPROBADOS_TOTAL > 0 ~ as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) ))
%>%
mutate(ttdistancia_t =sum(case_when(TURNO=="DISTANCIA" ~
as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) )) %>%
mutate(ttdiurno_t =sum(case_when(TURNO=="DIURNO" ~
as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) )) %>%
mutate(ttnocturno_t =sum(case_when(TURNO=="NOCTURNO" ~
as.numeric(REPROBADOS_TOTAL), TRUE ~ 0) )) %>%
mutate(tthombres = (REPROBADOS.MASCULINOS) %>%
mutate(ttmujeres= REPROBADOS.FEMENINOS) %>%
mutate(tt= sum(TOTAL.DE.MATRICULA.FINAL)) %>%
mutate(reprobados= (REPROBADOS_TOTAL) %>%
mutate(pct_reprobados = (reprobados/ tt)*100) %>%
mutate(pct_hombres = (( tthombres /reprobados)*100) /tt) %>%
mutate(pct_mujeres= (( ttmujeres/reprobados)*100)/ tt) %>%
mutate(pct_diurno= ((ttdiurno/ reprobados)*100) / tt) %>%
mutate(pct_nocturno = (( ttnocturno /reprobados)*100) /tt) %>%
mutate(pct_distancia= (( ttdistancia/reprobados)*100) /tt) %>%
mutate(pct_hombres_tt = (( tthombres/tt)*100) ) %>%
mutate(pct_mujeres_tt= (( ttmujeres/tt)*100) ) %>%
mutate(pct_diurno_tt= (( ttdiurno/tt)*100) ) %>%
mutate(pct_nocturno_tt = (( ttnocturno/tt)*100) ) %>%
mutate(pct_distancia_tt= (( ttdistancia/tt)*100) ) %>%
mutate(posee_nocturno = sum(case_when(TOTAL.DE.MATRICULA.FINAL > 0
& TURNO == "NOCTURNO"~ 1, TRUE ~ 0))) %>%
mutate(posee_diurno = sum(case_when(TOTAL.DE.MATRICULA.FINAL > 0 &
TURNO == "DIURNO"~ 1, TRUE ~ 0))) %>%
mutate(posee_distancia = sum(case_when(TOTAL.DE.MATRICULA.FINAL > 0
& TURNO == "DISTANCIA"~ 1, TRUE ~ 0))) %>%
mutate(posee_hombre = (case_when(TOTAL.DE.MATRICULA.FINAL > 0 &
REPROBADOS.MASCULINOS > 0 ~ 1, TRUE ~ 0))) %>%
mutate(posee_mujer = (case_when(TOTAL.DE.MATRICULA.FINAL > 0 &
REPROBADOS.FEMENINOS > 0 ~ 1, TRUE ~ 0))) %>%

```

```

select(tt,pct_reprobados, reprobados, tthombres, ttmujeeres,
pct_hombres, pct_mujeres,
ttdistancia,ttnocturno,ttdiurno,pct_distancia,pct_nocturno,pct_diurno
,pct_hombres_tt,
pct_mujeres_tt,pct_distancia_tt,pct_nocturno_tt,pct_diurno_tt,
posee_nocturno, posee_diurno,
posee_distancia,posee_hombre,posee_mujer) %>%
unique %>%
ungroup(MUNICIPIO, CÓDIGO.DE.MUNICIPIO) %>%
group_by(MUNICIPIO, CÓDIGO.DE.MUNICIPIO) %>%
mutate(pct_reprobados2= sum(pct_reprobados))
summarize(sum(tt),sum(pct_reprobados))#, reprobados, tthombres,
ttmujeeres, pct_hombres, pct_mujeres,
ttdistancia,ttnocturno,ttdiurno,pct_distancia,pct_nocturno,pct_diurno
,pct_hombres_tt,
pct_mujeres_tt,pct_distancia_tt,pct_nocturno_tt,pct_diurno_tt,
posee_nocturno, posee_diurno,
posee_distancia,posee_hombre,posee_mujer) %>%

```

```

" mutate(posee_nocturno = (case_when(posee_nocturno > 0 ~ 1, TRUE ~
0))) %>%
mutate(posee_diurno= (case_when(posee_diurno > 0 ~ 1, TRUE ~ 0)))
%>%
mutate(posee_distancia = (case_when(posee_distancia > 0~ 1 , TRUE ~
0))) %>%
mutate(posee_hombre = (case_when(posee_hombre > 0 ~ 1, TRUE ~ 0)))
%>%
mutate(posee_mujer = (case_when(posee_mujer > 0 ~ 1, TRUE ~ 0)))
%>%
select(tt,pct_reprobados, reprobados, tthombres, ttmujeeres,
pct_hombres, pct_mujeres,
ttdistancia,ttnocturno,ttdiurno,pct_distancia,pct_nocturno,pct_diurno
,pct_hombres_tt,
pct_mujeres_tt,pct_distancia_tt,pct_nocturno_tt,pct_diurno_tt,
posee_nocturno, posee_diurno,
posee_distancia,posee_hombre,posee_mujer) %>%
unique
"

```

```
#####
#####
str(datos_temp2)
datos_temp2 %>%
  filter(posee_diurno>0)

write.csv(data.frame(datos_temp2),"pct_reprobados_municipio_sexo_turno_2015.csv")

data_perct = read.csv("pct_reprobados_municipio_sexo_turno_2015.csv")
str(data_perct)

#correlacion
#Alta correlación entre los niveles de reprobados con la cantidad de
alumnos reprobados del sexo masculino
plot(data_perct$pct_reprobados, data_perct$pct_hombres_tt)
#Alta correlación entre los niveles de reprobados con la cantidad de
alumnos reprobados del sexo femenino pero hay mas variabilidad
plot(data_perct$pct_reprobados, data_perct$pct_mujeres_tt)
#No parece existir correlación lineal aparente entre los niveles de
reprobados con la cantidad de alumnos reprobados y los alumnos
reprobados
#En el turno de distancia

plot(data_perct$pct_reprobados, data_perct$pct_distancia_tt)

#No parece existir correlación lineal aparente entre los niveles de
reprobados con la cantidad de alumnos reprobados y los alumnos
reprobados
#En el turno de distancia

plot(data_perct$pct_reprobados, data_perct$pct_nocturno_tt)
#No parece existir correlación lineal aparente entre los niveles de
reprobados con la cantidad de alumnos reprobados y los alumnos
reprobados
#En el turno nocturno

#Si parece existir correlación lineal aparente entre los niveles de
reprobados con la cantidad de alumnos reprobados y los alumnos
reprobados
#En el turno diurno
```

```

plot(data_perct$pct_reprobados, data_perct$pct_diurno_tt)

#Correlaciones

#Hombres
cor.test(data_perct$pct_reprobados, data_perct$pct_hombres_tt)

#mujeres
cor.test(data_perct$pct_reprobados, data_perct$pct_mujeres_tt)

#turno diurno
cor.test(data_perct$pct_reprobados, data_perct$pct_diurno_tt)

#25 mayores municipios con porcentaje de reprobados en general

muni_mayor = arrange(data_perct, desc(data_perct$pct_reprobados))
(muni_mayor_gp = head(muni_mayor,n=25))

write.csv(data.frame(datos_temp), "pct_reprobados_municipio_mayor_gene
ral.csv")

#25 menores municipios con porcentaje de reprobados en general

muni_min = arrange(data_perct, (data_perct$pct_reprobados))
muni_min_gp = head(muni_min,n=25)

#25 mayores municipios con porcentaje de reprobados en general
hombres

muni_mayor_hombres = arrange(data_perct,
desc(data_perct$pct_hombres_tt))
(muni_mayor_gp_hombres = head(muni_mayor_hombres,n=25))

```

```
#25 menores municipios con porcentaje de reprobados en general  
hombres
```

```
muni_min_hombres = arrange(data_perct, (data_perct$pct_hombres_tt))  
muni_min_gp_hombres = head(muni_min,n=25)
```

```
#25 mayores municipios con porcentaje de reprobados en general  
mujeres
```

```
muni_mayor_mujeres = arrange(data_perct,  
desc(data_perct$pct_mujeres_tt))  
muni_mayor_gp_mujeres = head(muni_mayor_mujeres,n=25)
```

```
#25 menores municipios con porcentaje de reprobados en general  
mujeres
```

```
muni_min_mujeres = arrange(data_perct, (data_perct$pct_mujeres_tt))  
muni_min_gp_mujeres = head(muni_min_mujeres,n=25)
```

```
table(data_perct$pct_nocturno_tt)
```

```
#25 mayores municipios con porcentaje de reprobados en general  
nocturno
```

```
muni_mayor_nocturno = arrange(data_perct,  
desc(data_perct$pct_nocturno_tt))  
(muni_mayor_gp_nocturno = head(muni_mayor_nocturno,n=25))
```

```
#25 menores municipios con porcentaje de reprobados en general  
nocturno
```

```
muni_min_nocturno = arrange(data_perct, (data_perct$pct_nocturno_tt))  
(muni_min_gp_nocturno = head(muni_min_mujeres,n=25))
```

```
#220 no tienen reprobados nocturnos averiguar cuantas escuelas
tienen este tipo de turno
```

```
#25 mayores municipios con porcentaje de reprobados en general
distancia
```

```
muni_mayor_distancia = arrange(data_perct,
desc(data_perct$pct_distancia_tt))
(muni_mayor_gp_distancia = head(muni_mayor_distancia,n=25))
table(data_perct$pct_nocturno_tt)
```

```
#25 menores municipios con porcentaje de reprobados en general
distancia
```

```
muni_min_distancia = arrange(data_perct,
(data_perct$pct_distancia_tt))
(muni_min_gp_distancia = head(muni_min_distancia,n=25))
```

```
table(data_perct$pct_distancia_tt)
```

```
muni_mayor_diurno = arrange(data_perct,
desc(data_perct$pct_diurno_tt))
(muni_mayor_gp_diurno = head(muni_mayor_diurno,n=25))
table(data_perct$pct_diurno_tt)
```

```
#25 menores municipios con porcentaje de reprobados en general
distancia
```

```
muni_min_diurno = arrange(data_perct, (data_perct$pct_diurno_tt))
(muni_min_gp_diurno = head(muni_min_diurno,n=25))
```

```
table(data_perct$pct_diurno_tt)
```

```
#install.packages("mgcv")
library(mgcv)
```

```

#Buscando relación geográfica entre las variables
#data=read.csv("base_luz_perca.csv")
#str(data)
#attach(data)
library(spgwr)
#install.packages("spgwr")
#install.packages("ggExtra")
#install.packages("spdep")
library(tidyverse)
library(ggExtra) # for awesome plots
library(rgdal)
library(spdep)
library(spgwr)
#install.packages("reshape")
#library
#install.packages("spatialreg")
library(spatialreg)
library(car)

#Análisis espacial de reprobados en general

ny <- readOGR("municipios_regresion.shp")
glimpse(ny@data)

data_perct = read.csv("pct_reprobados_municipio_sexo_turno_2015.csv")
str(data_perct)
data_perct$CÓDIGO.DE.MUNICIPIO=
as.character(data_perct$CÓDIGO.DE.MUNICIPIO)
censo= data_perct %>%
  mutate(COD_MUN4.C.254_qgis = case_when( nchar(CÓDIGO.DE.MUNICIPIO)
< 4 ~ paste0("0",CÓDIGO.DE.MUNICIPIO), TRUE ~ CÓDIGO.DE.MUNICIPIO))

censo$COD_MUN4.C.254_qgis= factor(censo$COD_MUN4.C.254_qgis)
str(censo$COD_MUN4.C.254_qgis)
str(ny@data$COD_MUN4)

ny@data= ny@data %>%
  inner_join(censo, by = c("COD_MUN4" = "COD_MUN4.C.254_qgis"))

#ny@data= ny@data %>%
  str(ny@data)

```

```

#filter()

#plot(ny)
#plot(ny[which(ny@data$COD_MUN4 == "0101"),], col="red")
#ssplot(ny)

#plot(ny, col = "lightgrey")# plot the london_sport
#objectsel <- ny[which(ny@data$COD_MUN4 == "1103"),]
#plot(ny[which(ny@data$COD_MUN4 = "1103"), ], col = "turquoise", add
= TRUE)
#25 mayores municipios con porcentaje de reprobados en general
#muni_mayor = arrange(ny@data, desc(ny@data$pct_reprobados))
#muni_mayor_gp = head(muni_mayor,n=25)
#plot(ny, col = "lightgrey")# plot the london_sport
#text(ny)
#for(i in 1:nrow(muni_mayor_gp)){

# plot(ny[which(ny@data$COD_MUN4 == muni_mayor_gp$COD_MUN4[i]), ],
col = "turquoise", add = TRUE, text=muni_mayor_gp$NOM_MUN[i])
#text(ny[which(ny@data$COD_MUN4 == muni_mayor_gp$COD_MUN4[i]), ])
#text(ny@data$COD_MUN4)
#ny@data$NOM_MUN
#}
#sd
sd(data_perct$pct_reprobados)
sd(data_perct$pct_hombres_tt)
sd(data_perct$pct_mujeres_tt)
sd(data_perct$pct_diurno_tt)
sd(data_perct$pct_nocturno_tt)
sd(data_perct$pct_distancia_tt)
#correlacion

plot(data_perct$pct_reprobados, data_perct$pct_hombres_tt)

plot(data_perct$pct_reprobados, data_perct$pct_mujeres_tt)

#En el turno de distancia
plot(data_perct$pct_reprobados, data_perct$pct_distancia_tt)

plot(data_perct$pct_reprobados, data_perct$pct_nocturno_tt)

```

```

#Alta correlaciòn parente entre los niveles de reprobados con la
cantidad de alumnos reprobados y los alumnos reprobados
#En el turno diurno

plot(data_perct$pct_reprobados, data_perct$pct_diurno_tt)

#Hombres
cor.test(data_perct$pct_reprobados, data_perct$pct_hombres_tt)

#mujeres
cor.test(data_perct$pct_reprobados, data_perct$pct_mujeres_tt)

#turno diurno
cor.test(data_perct$pct_reprobados, data_perct$pct_diurno_tt)

#install.packages("tmap")
library(tmap)
vignette("tmap-getstarted")
#ny@data$

#25 mayores municipios con porcentaje de reprobados en general

muni_mayor = arrange(data_perct, desc(data_perct$pct_reprobados))
(muni_mayor_gp = head(muni_mayor,n=25))

#25 menores municipios con porcentaje de reprobados en general

muni_min = arrange(data_perct, (data_perct$pct_reprobados))
muni_min_gp = head(muni_min,n=25)
#Mapas para las tablas de reprobados en las diferentes categorías
#General
qtm(shp = ny, fill = "pct_reprobados", fill.n = 5, fill.palette =
"Reds", text="NOM_MUN", text.size=0.41, fill.title = "Porcentaje de
alumnos reprobados ES- 2018", title = "Alumnos reprobados 2018")# not
shown

#25 mayores municipios con porcentaje de reprobados en general
hombres

```

```

muni_mayor_hombres = arrange(data_perct,
desc(data_perct$pct_hombres_tt))
(muni_mayor_gp_hombres = head(muni_mayor_hombres,n=25))

#25 menores municipios con porcentaje de reprobados en general
hombres

muni_min_hombres = arrange(data_perct, (data_perct$pct_hombres_tt))
(muni_min_gp_hombres = head(muni_min,n=25))

#Hombres
qtm(shp = ny, fill = "pct_hombres_tt", fill.n = 5, fill.palette =
"Blues", text="NOM_MUN", text.size=0.41, fill.title = "Porcentaje de
alumnos reprobados de sexo masculino ES -2018",title = "Alumnos
reprobados de sexo masculino 2018")# not shown

#25 mayores municipios con porcentaje de reprobados en general
mujeres

muni_mayor_mujeres = arrange(data_perct,
desc(data_perct$pct_mujeres_tt))
(muni_mayor_gp_mujeres = head(muni_mayor_mujeres,n=25))

#25 menores municipios con porcentaje de reprobados en general
mujeres

muni_min_mujeres = arrange(data_perct, (data_perct$pct_mujeres_tt))
(muni_min_gp_mujeres = head(muni_min_mujeres,n=25))

#Mujeres
qtm(shp = ny, fill = "pct_mujeres_tt", fill.n = 5, fill.palette =
"Blues", text="NOM_MUN", text.size=0.45, fill.title = "Porcentaje de

```

```
alumnos reprobados de sexo femenino ES -2018",title = "Alumnos
reprobados de sexo femenino ES -2018")# not shown
```

```
muni_mayor_diurno = arrange(data_perct,
desc(data_perct$pct_diurno_tt))
(muni_mayor_gp_diurno = head(muni_mayor_diurno,n=25))
table(data_perct$pct_diurno_tt)
```

```
#25 menores municipios con porcentaje de reprobados en general
distancia
```

```
muni_min_diurno = arrange(data_perct, (data_perct$pct_diurno_tt))
(muni_min_gp_diurno = head(muni_min_diurno,n=25))
```

```
#Diurno
```

```
qtm(shp = ny, fill = "pct_diurno_tt", fill.n = 5, fill.palette =
"Greens", text="NOM_MUN", text.size=0.45, fill.title = "Porcentaje de
alumnos reprobados del turno diurno ES -2018",title = "Alumnos
reprobados del turno diurno 2018")# not shown
```

```
#25 mayores municipios con porcentaje de reprobados en general
nocturno
```

```
muni_mayor_nocturno = arrange(data_perct,
desc(data_perct$pct_nocturno_tt))
(muni_mayor_gp_nocturno = head(muni_mayor_nocturno,n=25))
```

```
#25 menores municipios con porcentaje de reprobados en general
nocturno
```

```
muni_min_nocturno = arrange(data_perct, (data_perct$pct_nocturno_tt))
(muni_min_gp_nocturno = head(muni_min_mujeres,n=25))
```

```
#Nocturno
```

```
qtm(ny, symbols.size = "posee_nocturno", fill.palette
="Greens",fill = "pct_nocturno_tt",
```

```

    fill.title = "Porcentaje de alumnos reprobados del turno
Nocturno ES -2018",
    symbols.title.size= "Municipios que realizan turno
nocturno",title = "Alumnos reprobados del turno Nocturno 2018",
    symbols.id= "COD_MUN4",
    text="NOM_MUN",text.size=0.45, symbols.text.size=0.3)

#25 mayores municipios con porcentaje de reprobados en general
distancia

    muni_mayor_distancia = arrange(data_perct,
desc(data_perct$pct_distancia_tt))
    (muni_mayor_gp_distancia = head(muni_mayor_distancia,n=25))
    table(data_perct$pct_nocturno_tt)

#25 menores municipios con porcentaje de reprobados en general
distancia

    muni_min_distancia = arrange(data_perct,
(data_perct$pct_distancia_tt))
    (muni_min_gp_distancia = head(muni_min_distancia,n=25))

#Distancia
    qtm(ny, symbols.size = "posee_distancia", fill.palette
="Greens",fill = "pct_distancia_tt",
    fill.title = "Porcentaje de alumnos reprobados del turno
distancia ES -2018",title = "Alumnos reprobados del turno distancia
2018",
    symbols.title.size= "Municipios que realizan turno nocturno",
    symbols.id= "COD_MUN4",
    text="NOM_MUN",text.size=0.45, symbols.text.size=0.3)

#Autocorrelacion espacial de la variable

    id <- row.names(as(ny, "data.frame"))
    coords <- coordinates(ny)
    ny_nb <- knn2nb(knearneigh(coords, k = 4), row.names = id)

```

```

plot(ny, main="Vecindarios")
plot(ny_nb, coords, add=T, main = "KNN = 4 Neighborhood")
#Se observa evidencia de autocorrelación global al 95% de confianza
moran.test(ny@data$pct_reprobados, nb2listw(ny_nb, style="B"))
#Se observa evidencia de autocorrelación global al 95% de confianza
moran.test(ny@data$pct_hombres_tt, nb2listw(ny_nb, style="B"))
#No se observa evidencia de autocorrelación global al 95% de
confianza
moran.test(ny@data$pct_mujeres_tt, nb2listw(ny_nb, style="B"))
ny_nb$weights

#probando moran test en hombres con diferentes estilos
moran.test(ny@data$pct_hombres_tt, nb2listw(ny_nb, style="W"))
moran.test(ny@data$pct_hombres_tt, nb2listw(ny_nb, style="S"))
moran.test(ny@data$pct_hombres_tt, nb2listw(ny_nb, style="C"))

#turnos
#No se observa evidencia de autocorrelación global al 95% de
confianza
moran.test(ny@data$pct_distancia_tt, nb2listw(ny_nb, style="B"))
#Se observa evidencia de autocorrelación global al 95% de confianza
moran.test(ny@data$pct_diurno_tt, nb2listw(ny_nb, style="B"))
#No se observa evidencia de autocorrelación global al 95% de
confianza
    moran.test(ny@data$pct_nocturno_tt, nb2listw(ny_nb, style="B"))

#probando getis ord

ny@data$rep.Gi= localG(ny@data$pct_reprobados,
listw=nb2listw(ny_nb, style="W"),zero.policy = TRUE)
ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Municipios - Alumnos reprobados",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)
ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

ny@data$rep.Gi= localG(ny@data$pct_hombres_tt,
listw=nb2listw(ny_nb, style="W"),zero.policy = TRUE)

```

```

ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Reprobados Sexo masculino",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)
ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

```

```

ny@data$rep.Gi= localG(ny@data$pct_mujeres_tt,
listw=nb2listw(ny_nb, style="W"),zero.policy = TRUE)
ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Reprobados Sexo Femenino",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)
ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

```

```

ny@data$rep.Gi= localG(ny@data$pct_diurno_tt, listw=nb2listw(ny_nb,
style="W"),zero.policy = TRUE)
ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Reprobados Turno diurno",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)
ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

```

```

ny@data$rep.Gi= localG(ny@data$pct_nocturno_tt,
listw=nb2listw(ny_nb, style="W"),zero.policy = TRUE)
ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Reprobados Turno nocturno",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)

```

```

ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

ny@data$rep.Gi= localG(ny@data$pct_distancia_tt,
listw=nb2listw(ny_nb, style="W"),zero.policy = TRUE)
ny@data$rep.Gi[is.na(ny@data$rep.Gi)] = 0
plot(ny, main="Local Getis-Ord Reprobados Turno distancia",
cex.main=1.0, font.main=1, axes=TRUE)
zkrit10 = qnorm(0.975) #confidence level
ny@data$rep.GisignH = ny@data$rep.Gi>zkrit10
plot(ny[ ny@data$rep.GisignH, ], col="red", add=TRUE)
ny@data$rep.GisignL = -zkrit10>ny@data$rep.Gi
plot(ny[ ny@data$rep.GisignL, ], col="BLUE", add=TRUE)

data_pib = read.csv("pib2015_mun.csv")
str(data_pib)
data_pib$CODIGO= as.character(data_pib$CODIGO)

censo_pib= data_pib%>%
  mutate(COD_MUN4.PIB = case_when( nchar(CODIGO) < 4 ~
paste0("0",CODIGO), TRUE ~ CODIGO))

censo_pib$COD_MUN4.PIB= factor(censo_pib$COD_MUN4.PIB)
str(censo_pib)

ny@data= ny@data %>%
  inner_join(censo_pib, by = c("COD_MUN4" = "COD_MUN4.PIB"))

#Regresiones

plot(ny@data$pct_reprobados ~ ny@data$PIBPC_2015)
plot(ny@data$pct_reprobados ~ ny@data$CEPC_2015)
hist(ny@data$PIBPC_2015)
boxplot(ny@data$PIBPC_2015)
ny.ols <- lm(ny@data$pct_reprobados ~ ny@data$PIBPC_2015, data =
ny)
summary(ny.ols)
#Min. 1st Qu. Median Mean 3rd Qu. Max.
#44.7 1095.8 2686.3 16174.0 8328.2 329235.6
#El gran porcentaje de municipios poseen un PIB pequeño, mediana
2686.3
#Es recomendable categorizar la variable PIB con método de árbol

```

```

#####
#árbol regresión
library(rpart)
(tmod <- rpart(ny@data$pct_reprobados ~ ny@data$PIBPC_2015 , ny))

summary(tmod)
plot(tmod)
text(tmod)
tmod

printcp(tmod)

ny@data = ny@data %>%
  mutate(tree_r = case_when(PIBPC_2015< 381.19 ~ 1,
                             PIBPC_2015>=381.19 &
PIBPC_2015<692.7386 ~ 2,
                             PIBPC_2015>=692.7386 &
PIBPC_2015<1060.156 ~ 3,
                             PIBPC_2015>=1060.156 &
PIBPC_2015<1193.496 ~ 4,
                             PIBPC_2015>=1193.496 &
PIBPC_2015<2481.467 ~ 5,
                             PIBPC_2015>=2481.467 &
PIBPC_2015<4050.064 ~ 6,
                             PIBPC_2015>=4050.064 &
PIBPC_2015<5073.224 ~ 7,
                             PIBPC_2015>=5073.224 &
PIBPC_2015<6049.284 ~ 8,
                             PIBPC_2015>=6049.284 &
PIBPC_2015<6049.284 ~ 9,
                             PIBPC_2015>=6049.284 ~ 10
                             )
                             )

bestcp <- tmod$cpstable[which.min(tmod$cpstable[, "xerror"]), "CP"]

summary(residuals(tmod))
plot(predict(tmod), residuals(tmod))

library(agricolae)

```

```

library(lmtest)
library(nortest)

ny.ols_tree <- lm(ny@data$pct_reprobados ~ factor(ny@data$tree_r),
data = ny)
summary(ny.ols_tree)

plot(ny.ols_tree)

naov = aov(ny@data$pct_reprobados ~ factor(ny@data$tree_r))
summary(naov)

#Valor de R2 es muy pequeño

#prueba de supuestos
#Normalidad
shapiro.test(residuals(ny.ols_tree))
#no existe normalidad, verificar outliers
ks.test(ny.ols_tree$residuals, pnorm, mean(ny.ols_tree$residuals),
sd(ny.ols_tree$residuals))
#si existe normalidad
ad.test(ny.ols_tree$residuals)
#no existe normalidad

#homocedasticidad
car::ncvTest(ny.ols_tree)
#homocedasticidad es cercana al 0.05 se recomienda comparar con
otro test
bartlett.test(ny.ols_tree$residuals ~ factor(ny@data$tree_r))
#si existe homocedasticidad
leveneTest(ny.ols_tree$residuals, factor(ny@data$tree_r),
center="median")
#Si existe homocedasticidad

#Independencia no geográfica
# Independencia
# -----
plot(ny.ols_tree$residuals)
dwtest(ny.ols_tree)
#no existe dependencia entre las variables - no considera vinculo
geográfico

```

```

table(factor(ny@data$tree_r))

summary(lm(naov))

plot(ny@data$pct_reprobados ~ factor(ny@data$tree_r))
plot(ny.ols_tree)
"
Residuals:
      Min       1Q   Median       3Q      Max
-2.7532 -1.0278 -0.1746  0.7127  6.4323

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.6924     0.2726   6.209 2.17e-09 ***
factor(ny@data$tree_r)2  1.0608     0.3973   2.670 0.008078 **
factor(ny@data$tree_r)3  1.6056     0.4781   3.358 0.000905 ***
factor(ny@data$tree_r)4  0.6055     0.5066   1.195 0.233098
factor(ny@data$tree_r)5  1.1285     0.3494   3.230 0.001401 **
factor(ny@data$tree_r)6  0.7732     0.3757   2.058 0.040599 *
factor(ny@data$tree_r)7  1.8693     0.4385   4.263 2.85e-05 ***
factor(ny@data$tree_r)8  0.6579     0.5243   1.255 0.210707
factor(ny@data$tree_r)10 0.8573     0.3116   2.751 0.006363 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.416 on 253 degrees of freedom
Multiple R-squared:  0.08855,    Adjusted R-squared:  0.05973
F-statistic: 3.072 on 8 and 253 DF,  p-value: 0.00254"

#####
#####3

ny.ols <- lm(ny@data$pct_reprobados ~ (ny@data$PIBPC_2015), data =
ny)
summary(ny.ols)
plot(ny.ols)
plot(ny@data$pct_reprobados ~ (ny@data$PIBPC_2015))
#no cumple supuestos

ny.olsh <- lm(ny@data$pct_hombres_tt ~ (ny@data$PIBPC_2015), data =
ny)

```

```

summary(ny.olsh)
#bajo R2

ny.olsm <- lm(ny@data$pct_mujeres_tt ~ (ny@data$PIBPC_2015), data =
ny)
summary(ny.olsm)
#bajo R2

ny.olsd <- lm(ny@data$pct_diurno_tt ~ (ny@data$PIBPC_2015), data =
ny)
summary(ny.olsd)
#bajo R2

ny.olsn <- lm(ny@data$pct_nocturno_tt ~ (ny@data$PIBPC_2015), data =
ny)
summary(ny.olsn)
#bajo R2

#####

ny.olsh.tree <- lm(ny@data$pct_hombres_tt ~ factor(ny@data$tree_r),
data = ny)
summary(ny.olsh.tree)
#bajo R2

ny.olsm <- lm(ny@data$pct_mujeres_tt ~ factor(ny@data$tree_r), data =
ny)
summary(ny.olsm)
#bajo R2

ny.olsd <- lm(ny@data$pct_diurno_tt ~ factor(ny@data$tree_r), data =
ny)
summary(ny.olsd)
#bajo R2

ny.olsn <- lm(ny@data$pct_nocturno_tt ~ factor(ny@data$tree_r), data
= ny)
summary(ny.olsn)
#bajo R2

ny.olsn <- lm(ny@data$pct_distancia_tt ~ factor(ny@data$tree_r), data
= ny)

```

```

summary(ny.olsn)

#####
#version robusta
library(rrcov)

LtsRobusta= ltsReg(ny@data$pct_reprobados ~ (ny@data$PIBPC_2015))
summary(LtsRobusta)
abline(LtsRobusta$coefficients, col="red")

par(mfrow=c(2,2))
plot(LtsRobusta)

#####
#versión espacial
ny_w <- nb2listw(ny_nb, style = "B", zero.policy = T)
lm.morantest(ny.ols, ny_w)
#Existe autocorrelación espacial

ny.sem <- errorsarlm(formula = ny@data$pct_reprobados ~
(ny@data$PIBPC_2015), data = ny, listw = ny_w, quiet =
T,method="eigen")
summary(ny.sem)

AIC(ny.ols,ny.sem)
#El modelo no mejora demasiado

library(spatialreg)
ny.lag <- lagsarlm(formula = ny@data$pct_reprobados ~
(ny@data$PIBPC_2015), data = ny, ny_w, zero.policy = T)
summary(ny.lag)

AIC(ny.ols,ny.sem,ny.lag)
#No hay de residual autocorrelation entre las variables

#####
#

ny.ols_tree_glm <- glm(ny@data$pct_reprobados ~
factor(ny@data$tree_r), data = ny)
summary(ny.ols_tree_glm)

```

```

AIC(ny.ols,ny.ols_tree, ny.ols_tree_glm)

ny.ols_tree_gam <- gam(ny@data$pct_reprobados ~
factor(ny@data$tree_r), data = ny)
summary(ny.ols_tree_gam)

AIC(ny.ols,ny.ols_tree, ny.ols_tree_glm,  ny.ols_tree_gam)

#####
#####333

ny.ols_glm <- glm(ny@data$pct_reprobados ~ ny@data$PIBPC_2015, data =
ny)
summary(ny.ols_glm)

AIC(ny.ols,ny.ols_tree, ny.ols_tree_glm)

ny.ols_gam <- gam(ny@data$pct_reprobados ~ ny@data$PIBPC_2015, data =
ny)
summary(ny.ols_gam)

AIC(ny.ols,ny.ols_tree, ny.ols_tree_glm,  ny.ols_tree_gam)

#####
#####333

ny.ols_gam.tree <- gam((ny@data$pct_reprobados) ~
s(ny@data$PIBPC_2015), data = ny)
summary(ny.ols_gam.tree)

#####

#versión robusta
LtsRobustah= ltsReg(ny@data$pct_hombres_tt ~ factor(ny@data$tree_r))
summary(LtsRobustah)

LtsRobustam= ltsReg(ny@data$pct_mujeres_tt ~ factor(ny@data$tree_r))
summary(LtsRobustam)

LtsRobustad= ltsReg(ny@data$pct_diurno_tt ~ factor(ny@data$tree_r))

```

```

summary(LtsRobustad)

LtsRobustan= ltsReg(ny@data$pct_nocturno_tt ~ factor(ny@data$tree_r))
summary(LtsRobustan)
plot(LtsRobustan)
#####
##
#ESPACIAL

ny.sem <- errorsarlm(formula = ny@data$pct_reprobados ~
(ny@data$PIBPC_2015), data = ny, listw = ny_w, quiet =
T,method="eigen")
summary(ny.sem)

y.olsh.tree <- lm(ny@data$pct_hombres_tt ~ factor(ny@data$tree_r),
data = ny)
summary(ny.olsh.tree)
lm.morantest(y.olsh.tree, ny_w)

ny.olsm <- lm(ny@data$pct_mujeres_tt ~ factor(ny@data$tree_r), data =
ny)
summary(ny.olsm)
lm.morantest(ny.olsm, ny_w)

#bajo R2

ny.olsd <- lm(ny@data$pct_diurno_tt ~ factor(ny@data$tree_r), data =
ny)
summary(ny.olsd)
lm.morantest(ny.olsd, ny_w)
#bajo R2

ny.olsn <- lm(ny@data$pct_nocturno_tt ~ factor(ny@data$tree_r), data
= ny)
summary(ny.olsn)
lm.morantest(ny.olsn, ny_w)
#bajo R2

ny.olsd <- lm(ny@data$pct_distancia_tt ~ factor(ny@data$tree_r), data
= ny)
summary(ny.olsd)
lm.morantest(ny.olsd, ny_w)

```

