Nome Completo dos Integrantes do Grupo:

TIAGO MARTINS
VILSON FERRREIRA

Estudo de Caso

Instruções para o Estudo de Caso

Desenvolver um estudo de caso (aplicado a sua área de interesse, pública ou privada) que utilize ferramentas de análise e exploração geográfica e/ou técnicas de Estatística Espacial para resolver algum problema de negócios.
O grupo deve utilizar dados geográficos públicos (sistemáticos), obrigatoriamente, e também privados (da empresa ou organização que será analisada), se pertinente e factível.

O trabalho deverá propor e realizar análises (plausíveis) que possam ser implantadas através das ferramentas discutidas ou apresentadas durante o curso (ArcView GIS, ArcGIS, GeoDA, Quantum GIS, R, etc). Deve incluir obrigatoriamente Análises envolvendo ferramentas Desktop GIS e também Estatística Espacial.

- Grupos de até 5 pessoas
- Entregas:

- 13/Outubro: Entrega do Pré-Projeto
- 18/Outubro: Discussão dos Trabalhos em grupo
- 01/Novembro: Apresentação dos Trabalhos em Grupo e entrega do Relatório Final (pode ser em formato de artigo)

Estudo de caso

O pré-projeto deve conter as intenções do grupo em realizar o trabalho, destacando relevância e importância do caso proposto. Além disso, mais diretamente, o pré-projeto deve apresentar o objetivo que se pretende atingir.

As bases de dados que pretendem ser utilizadas devem estar descritas, no maior nível de detalhe explorado pelo grupo.

Utilizem ESTE DOCUMENTO (Especificação Estudo de Caso.DOCX) como modelo para o documento de descrição do pré-projeto (Times New Roman 12, espaçamento simples entre linhas e antes e depois dos parágrafos, limites de margem conforme este documento) e dos relatórios preliminar e final.

O pré-projeto deve conter no mínimo 2 páginas. A avaliação do pré-projeto será um feedback para ajudar o grupo na realização do projeto em si. Por isso, não economizem na especificação – dúvidas sobre a condução do trabalho deverão constar no pré-projeto.
Não há limites de páginas para o relatório final.

1. INTRODUÇÃO

O ENEM (Exame Nacional do Ensino Médio) é atualmente a principal forma de acompanhar o desempenho dos alunos de educação básica, em especial o ensino médio, e considerado de grande sucesso devido sua aceitação em diversas esferas, tais como:

1. Atribuição de vagas e bolsas em universidades públicas e privadas.
2. Participações de alunos e privilégios em programas de governo para o ensino superior.
3. Estímulo à competitividade entre as instituições de ensino, tanto pública quanto privadas.
4. Estímulo à competitividade dos municípios e suas gestões da educação.

O estudo a seguir, tem o como objetivo compreender como os dados se comportam através de uma analise exploratória, para então definir e testar hipóteses que possam explicar como as média gerais das notas do exame são impactadas por outras variáveis, incluindo variáveis geográficas. As observações e analises feitas neste estudo estão subdividas em duas principais visões:

1- Pela ótica das instituições de ensino. Analisando o desempenho de cada escola.
2- Pela ótica de todos dos inscritos que realizaram a prova. Analisando a base completa de microdados.

2. DADOS – COLETA E TRATAMENTO

Os dados utilizados para este estudo foram coletados no **Portal Brasileiro de Dados Abertos** (dados.gov.br), estando atualmente disponível como **Microdados do ENEM 2014**, em forma zip/csv de 1.2GB.

Por serem bases pesadas, com mais de 8 milhões de registros, a estratégia adotada foi utilizar um SGDB (Sistema gerenciador de banco de dados) para que a manipulação dos dados fosse possível. Com a possibilidade de criação de índices e a realização de junções entre tabelas, a análise e tratamento dos dados ficam mais fáceis.

O primeiro passo foi entender quais planilhas estavam disponíveis e que tipo de variável cada uma delas possuía. Para tal, o dicionário de dados (Dicionário_Microdados_Enem_2014.xlsx) foi essencial.

Após a primeira análise, as duas principais planilhas escolhidas para o presente estudo foram MICRODADOS_ENEM_2014.csv e PLANILHA_ENEM_ESCOLA_2014.xlsx. A primeira contém os dados de todos os candidatos cadastrados para a realização do Enem. A segunda, o desempenho de cada uma das escolas participantes.

Com relação a planilha de microdados, é importante ressaltar que os candidatos que não compareceram a prova, também estavam contidos nesta tabela. Portanto, foi necessária a exclusão desses registros para que as análises não fossem prejudicadas. Para tal, a estratégia utilizada foi remover os registros cujas notas de todas as categorias, inclusive redação, estivessem zeradas.

De acordo com o **Portal INEP**, dois critérios são exigidos para que os resultados das instituições sejam divulgados na tabela de escolas:

1- Possuir pelo menos 10 (dez) alunos concluintes do ensino médio regular seriado participantes do Enem 2014; e

2- Possuir pelo menos 50% de alunos participantes do Enem 2014, de acordo com os dados do Censo Escolar 2014.

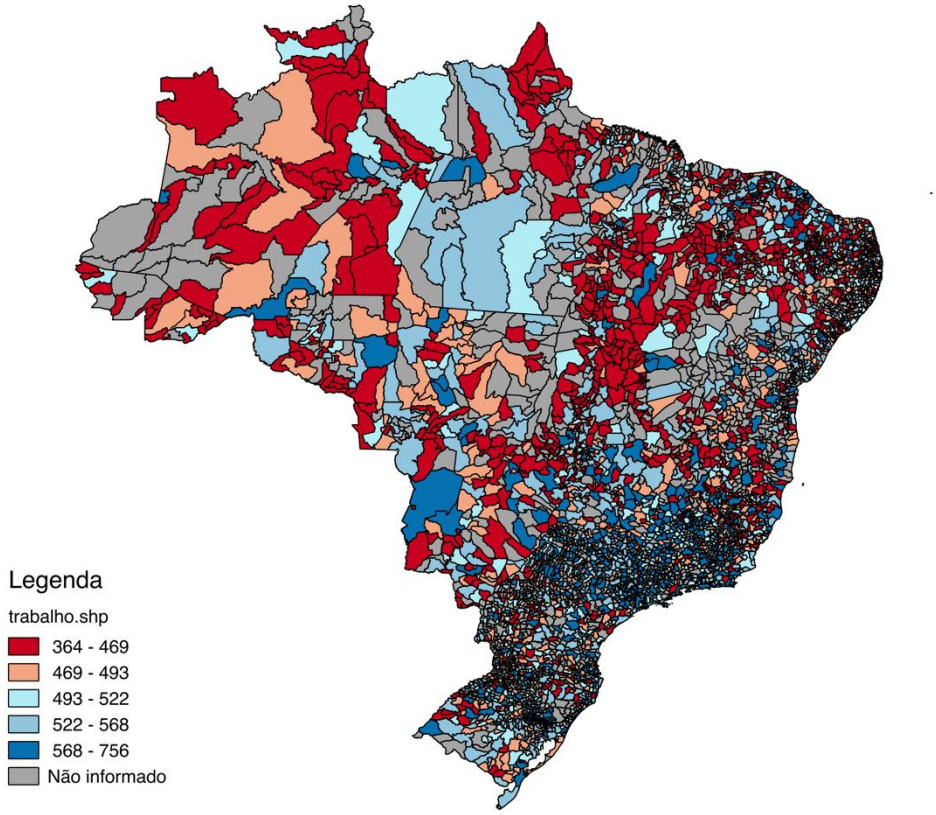
A tabela de escolas, por padrão, não possui o código IBGE do município, portanto foi realizada uma junção com a tabela de microdados através do campo cod_escola (constante nas duas tabelas), para que então ela fosse atualizada com esta variável.

3. ANÁLISE EXPLORATÓRIA – COMPORTAMENTO DOS DADOS

3.A – VISÃO POR ESCOLAS

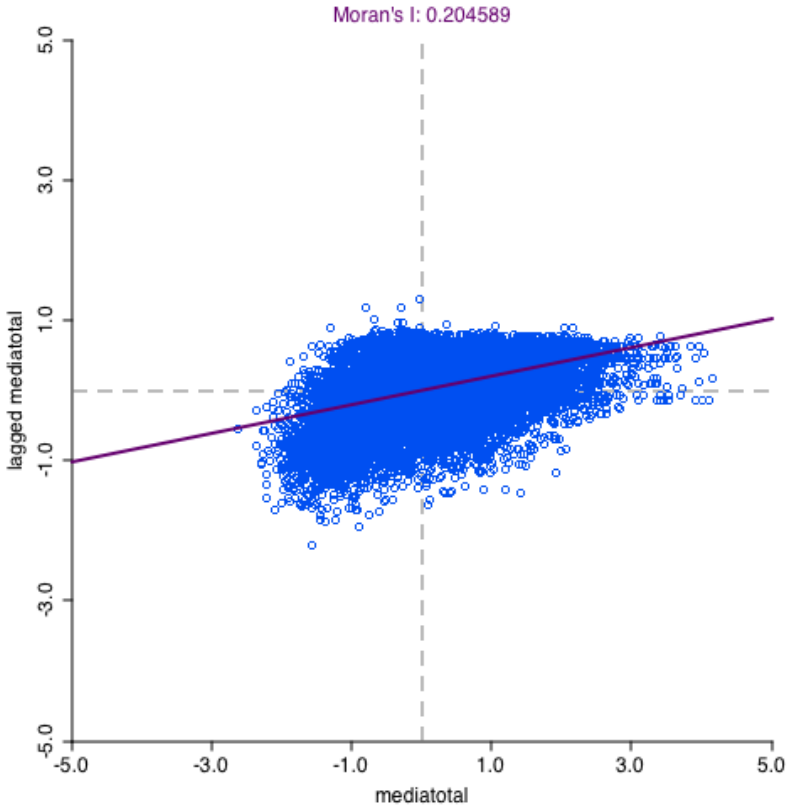
O primeiro passo, foi analisar como as notas médias das escolas estão distribuídas geograficamente por municípios. Para esta verificação, o software Quantum GIS foi utilizado. O shape de municípios do Brasil foi baixado em <http://www.codegeo.com.br/2013/04/shapefiles-do-brasil-para-download.html>. O mapa a seguir mostra uma média geral das notas das escolas através de suas regiões.

Imagem 1 –Notas por Municípios



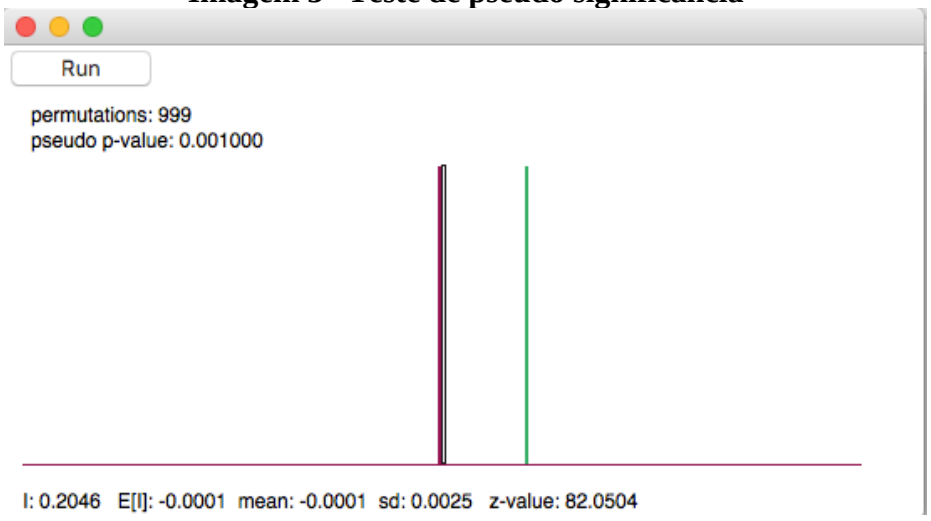
Analisando o mapa, é possível observar que na região Sul e Sudeste predominam notas mais altas, enquanto na região Nordeste e boa parte do Norte, concentram-se notas mais baixas.
Olhando apenas para o mapa, as regiões parecem ter alguma influência nas médias das notas. Para verificar esta hipótese, foi utilizado o software Geoda, e através da analise do índice de Moran (vizinhança por contiguidade - Queen) foi possível analisar se existem ou não, uma autocorrelação espacial.

Imagem 2 –Analise índice de Moran



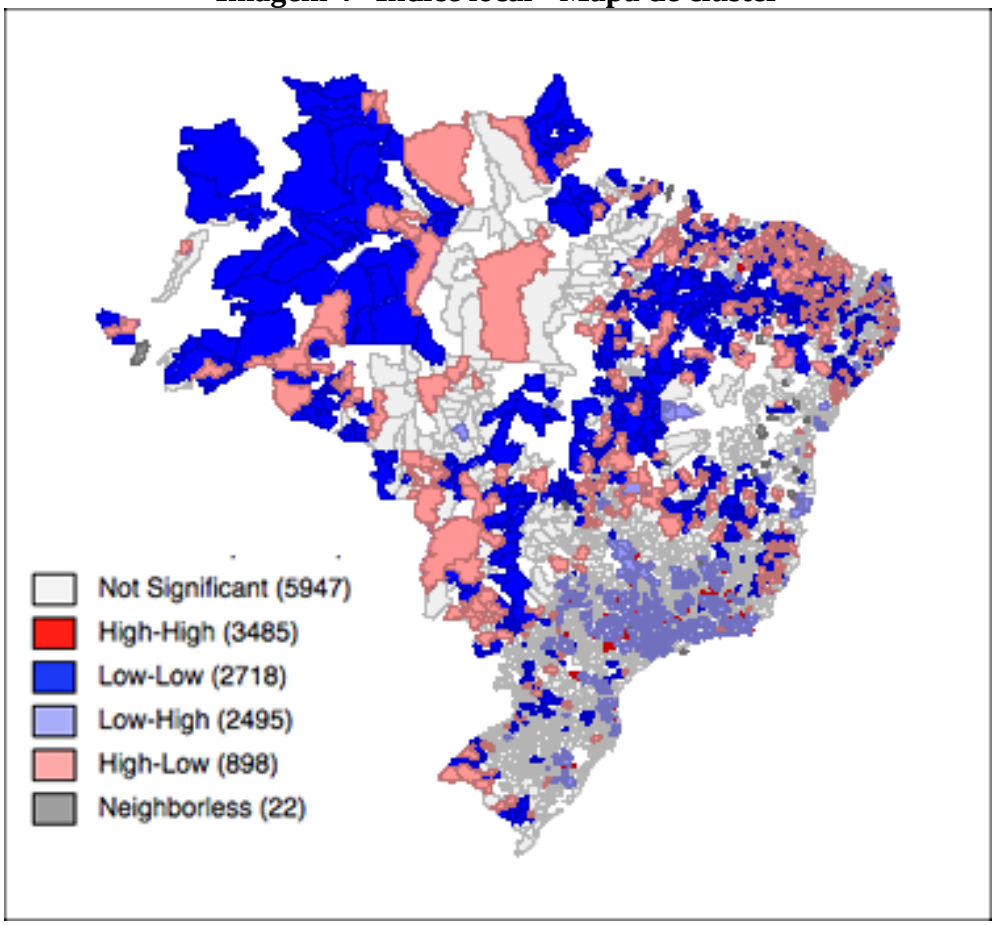
O índice de Moran apontou um valor de 0.204589. Para saber se este valor é significativo, foi necessário a realização de um teste de pseudo-significância. Neste caso, foram realizadas 999 permutações.

Imagem 3 –Teste de pseudo significância



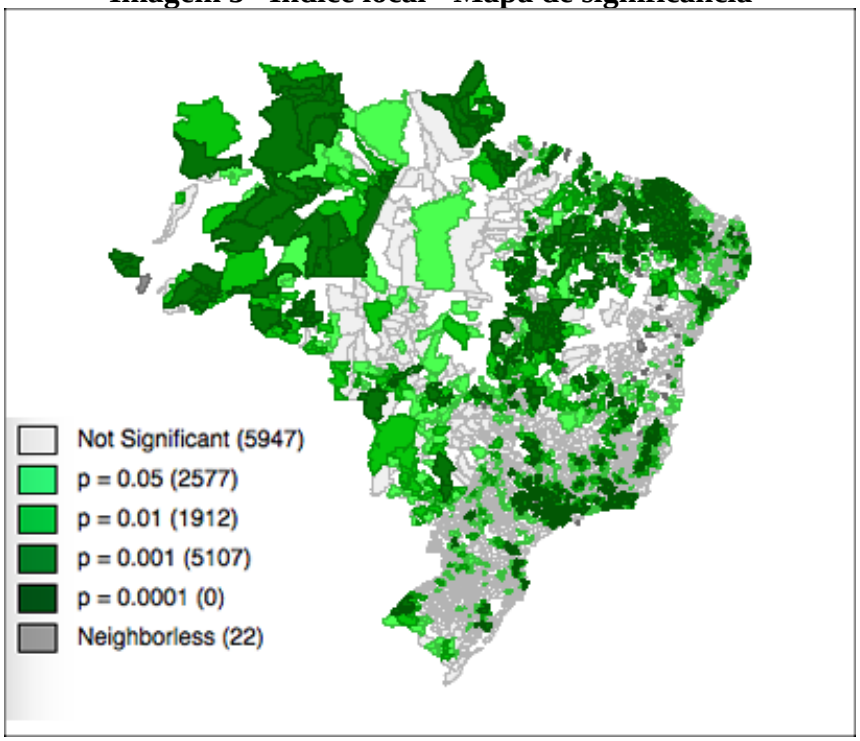
Analisando o p-valor, e observando que o índice calculado originalmente esta em uma das extremidades, a hipótese nula é rejeitada e conclui-se que, apesar de um valor aparentemente baixo, existe uma autocorrelação espacial.

Imagem 4 –Índice local - Mapa de cluster



No mapa acima (Imagem 4), é possível analisar a autocorrelação das notas entre os vizinhos: escolas com notas altas compartilhadas com vizinhos que também possuem notas altas, escolas com notas baixas com vizinhos que também possuem notas baixas e, por fim, escolas com vizinhos cujas notas são distintas. Abaixo (Imagem 5) é possível observar um mapa de significância dessas autocorrelações.

Imagem 5 –Índice local - Mapa de significância



Após atestada a autocorrelação espacial, o passo seguinte foi verificar como fica a representação das notas no mapa quando as escolas são analisadas por grupos de dependência administrativas, ou seja, as escolas são divididas entre municipais, estaduais, federais ou privadas.

Imagem 6 –Notas Escolas Federais

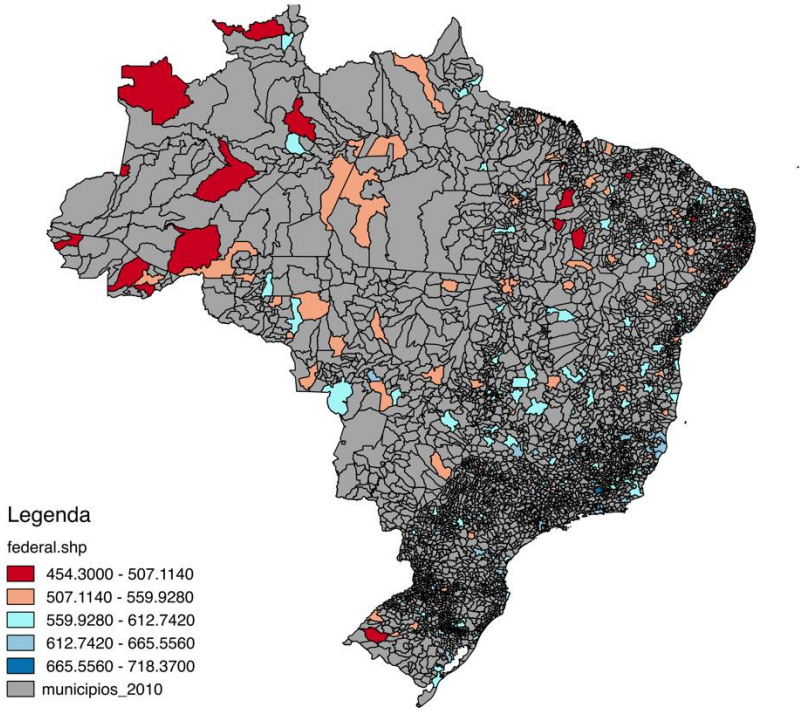


Imagem 7 –Notas Escolas Estaduais

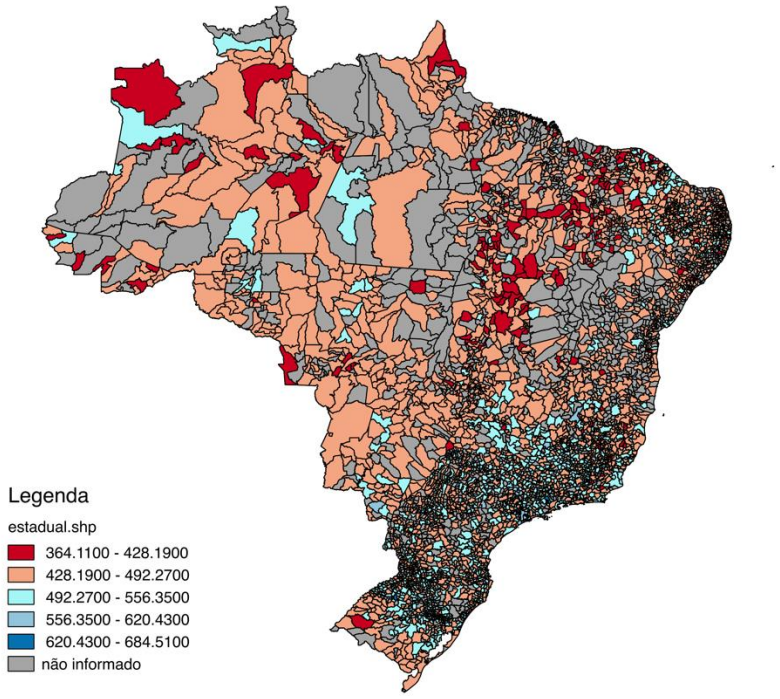


Imagem 8 – Notas Escolas Municipais

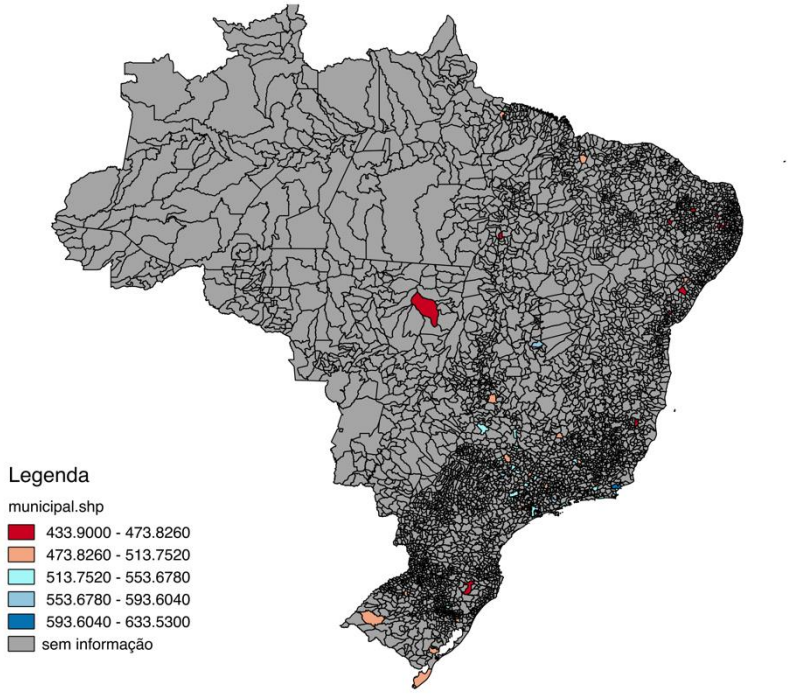
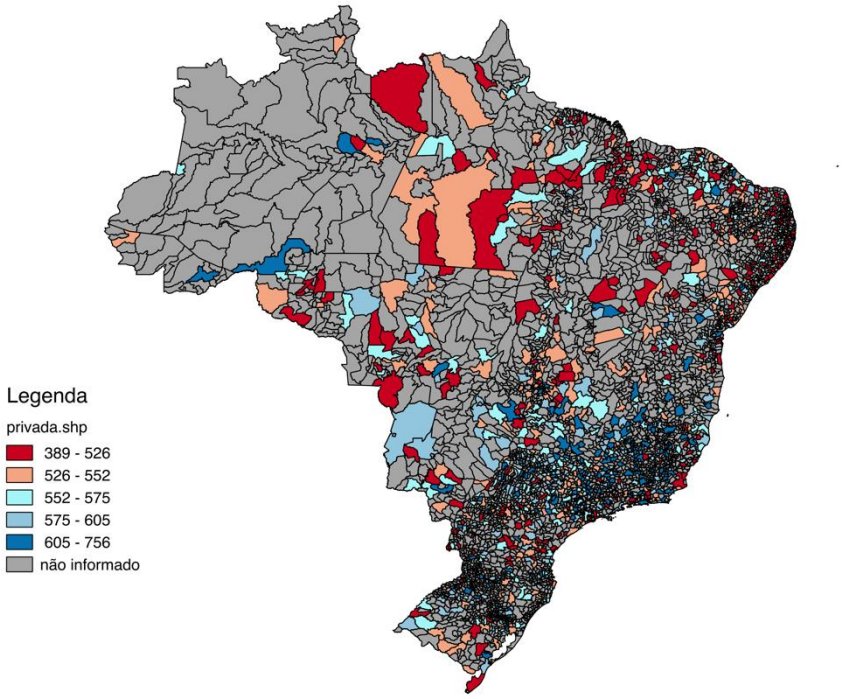


Imagem 9 – Notas Escolas Privadas



Através destes mapas, foi possível perceber uma tendência de notas mais baixas em instituições de ensino públicas, se comparadas com as instituições privadas.
Com o software Geoda, os mesmos shapes utilizados nos mapas acima foram importados e a partir daí, foram gerados alguns histogramas com as notas:

Imagem 10 – Histograma Escolas Federais

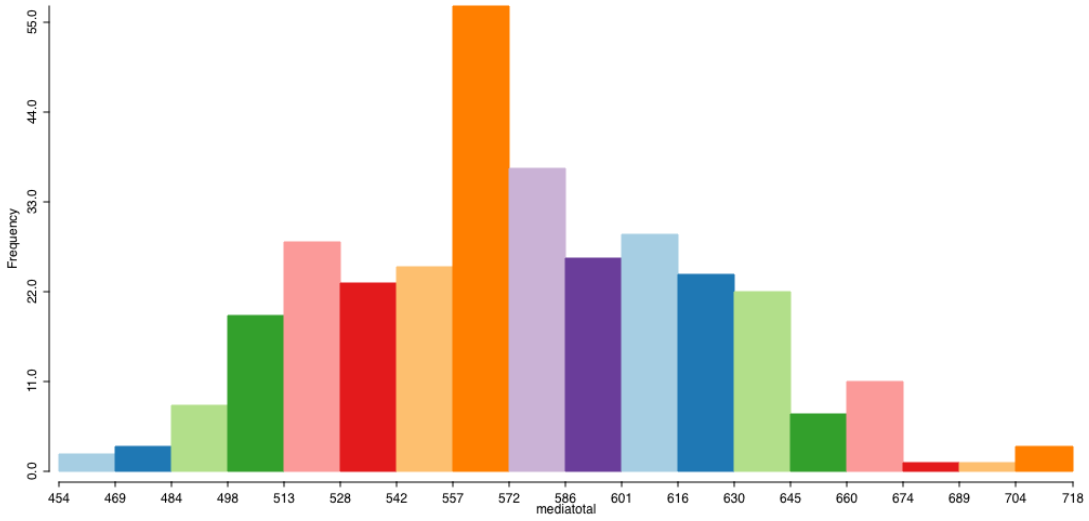


Imagem 11 – Histograma Escolas Estaduais

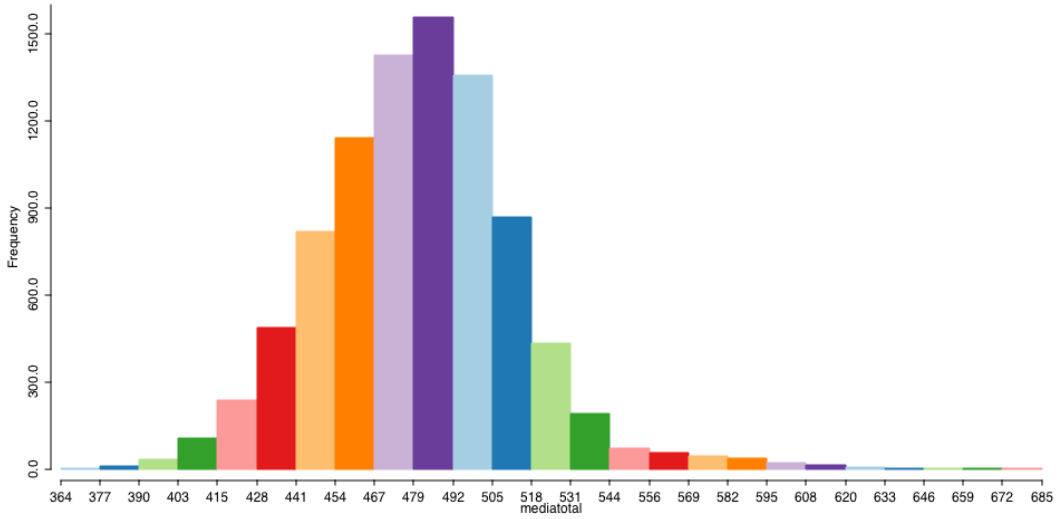


Imagem 12 – Histograma Escolas Municipais

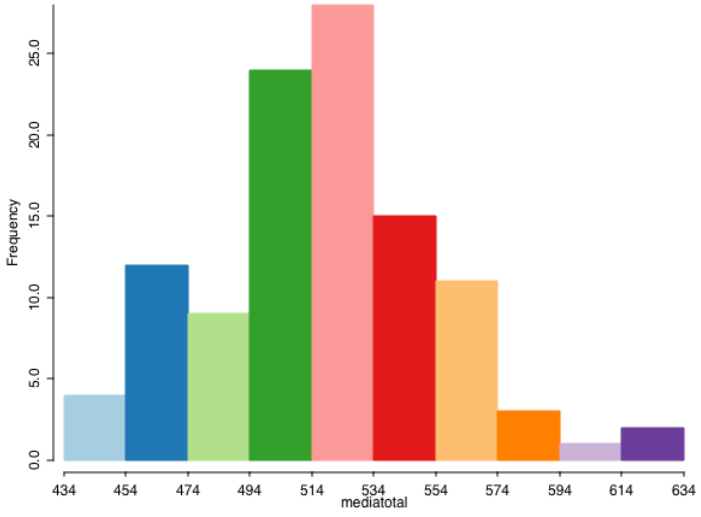
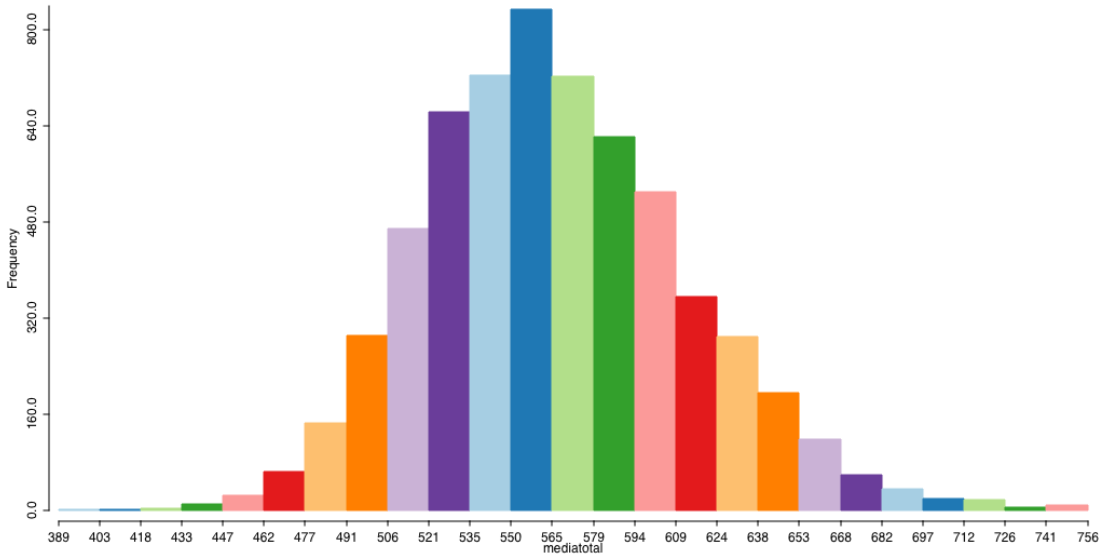


Imagem 13 – Histograma Escolas Privadas



3.B – VISÃO POR INDIVÍDUOS

Na análise por indivíduos foram elencadas 5 variáveis para serem analisadas geograficamente, baseado nos inscritos que realizaram a prova do ENEM, sendo elas:

- **Média Geral** baseada no município dos indivíduos a partir do dado COD_MUNICIPIO_RESIDENCIA.
- **Média** baseada no sexo, a partir do dado TP_SEXO sendo separadamente **Masculino** e **Feminino**.
- **Média** baseada no tipo de escola que o inscrito estuda (ou), a partir do dado TP_ESCOLA, sendo separadamente **Pública** ou **Privada**.

Com estas variáveis, torna-se possível avaliar o desempenho dos inscritos regionalmente, com diferenciação do desempenho pelo sexo, escola pública, escola privada. Além disso, ajuda a buscar detectar qual o reflexo deste desempenho e sua influência na Média do ENEM.
Como primeiro trabalho, foi avaliada a distribuição das médias de forma geral calculadas por municípios.

Imagem 14 – Box-plot da média geral por municípios

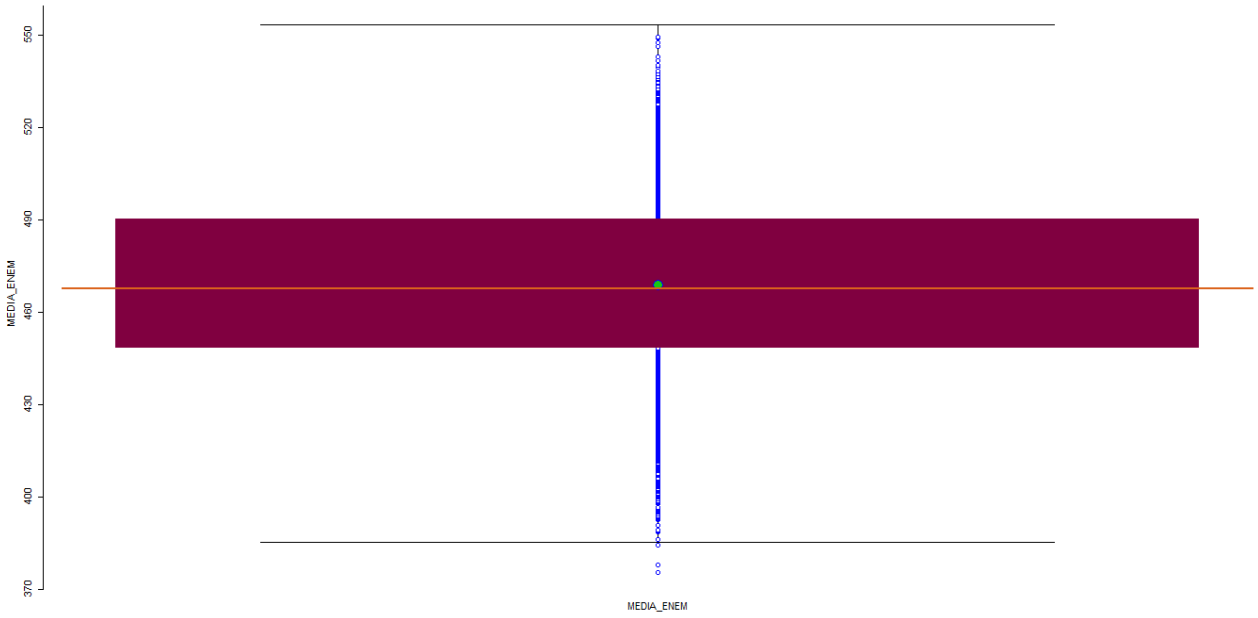
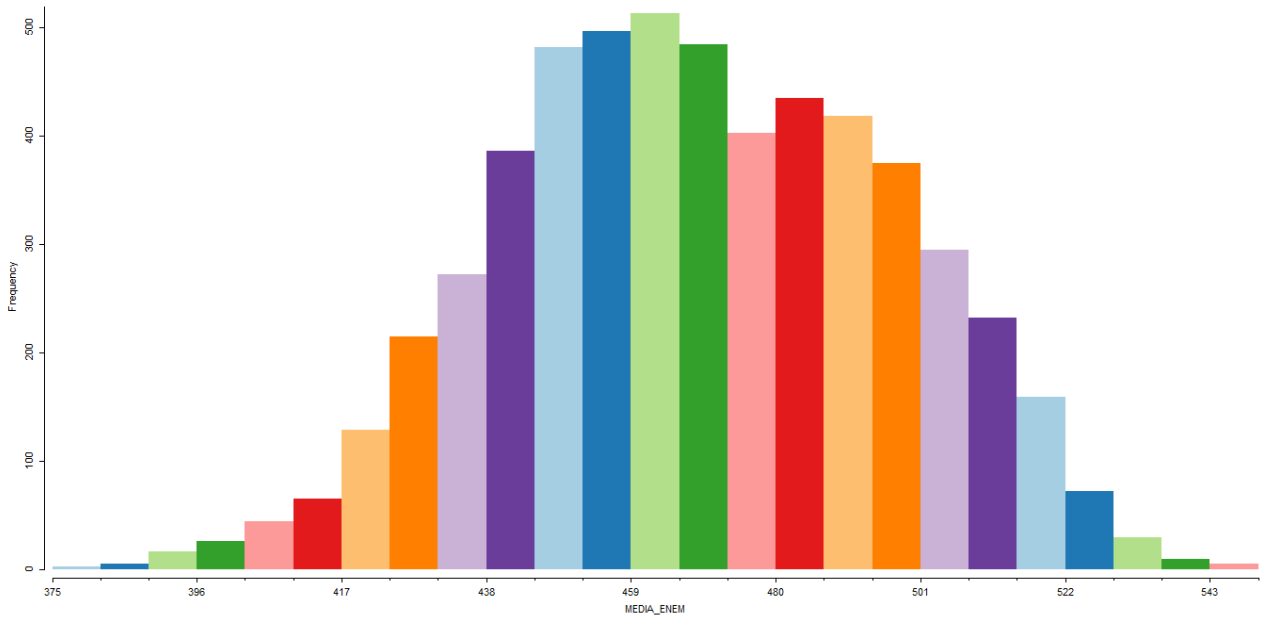


Imagem 15 – Histograma da média geral por município



A mesma análise foi realizada para o perfil de inscritos que estudaram em escola pública e privada. Desta forma começa-se a visualizar o quanto as médias daqueles que estudaram em escolas privadas são mais concentradas e de uma faixa maior, em relação àqueles que estudaram em escolas públicas.

Imagem 16 – Box-plot da média por estudo em Escola Pública

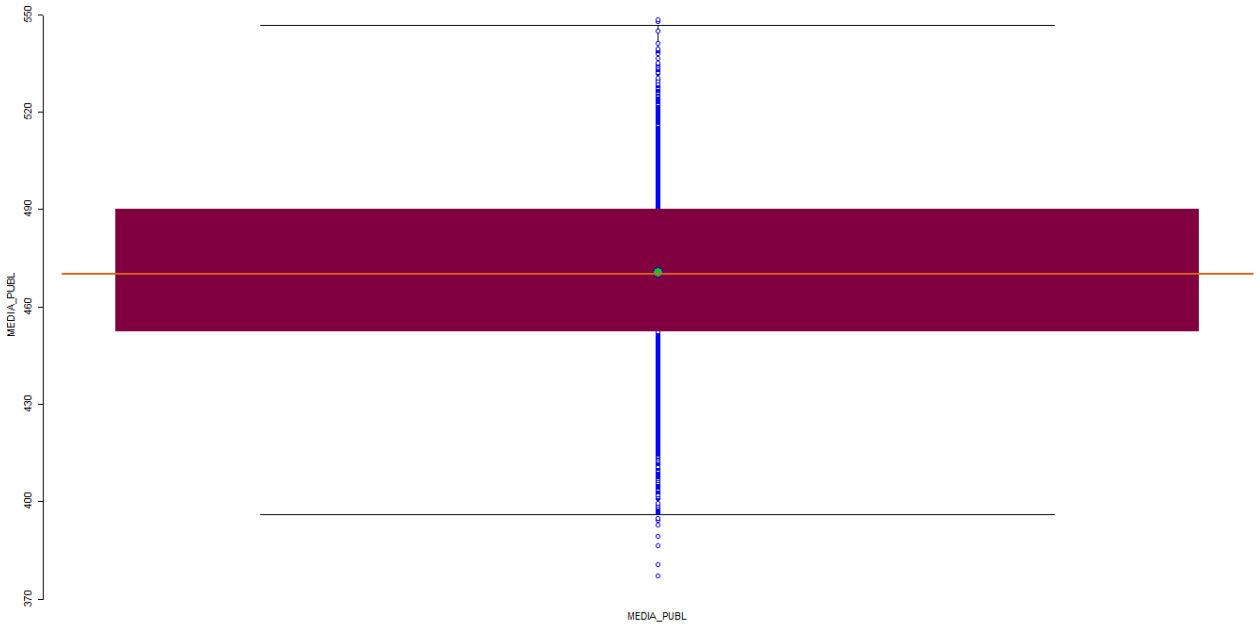


Imagem 17 – Histograma da média por estudo em Escola Pública

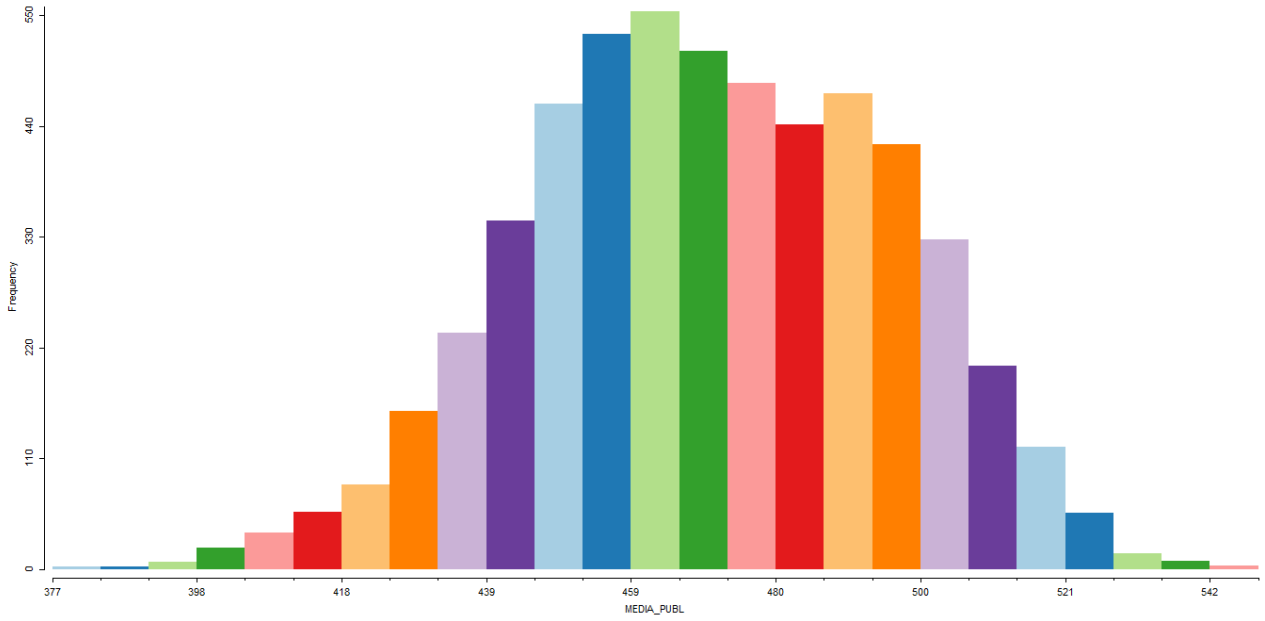


Imagem 18 – Box-plot da média por estudo em Escola Privada

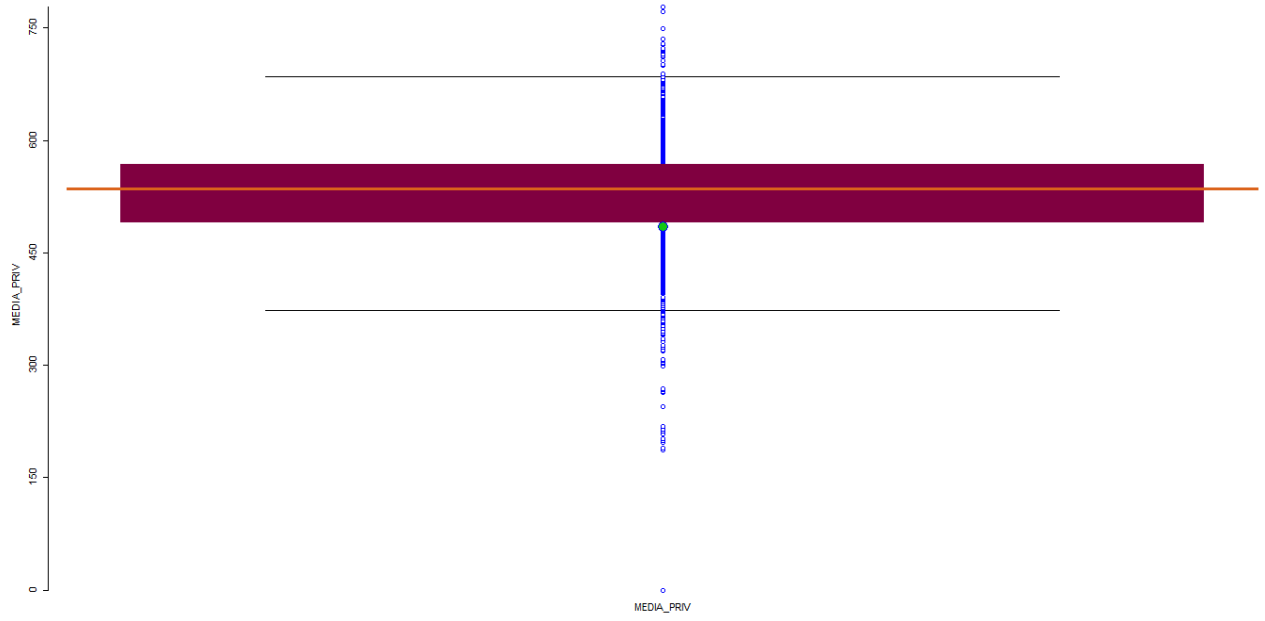
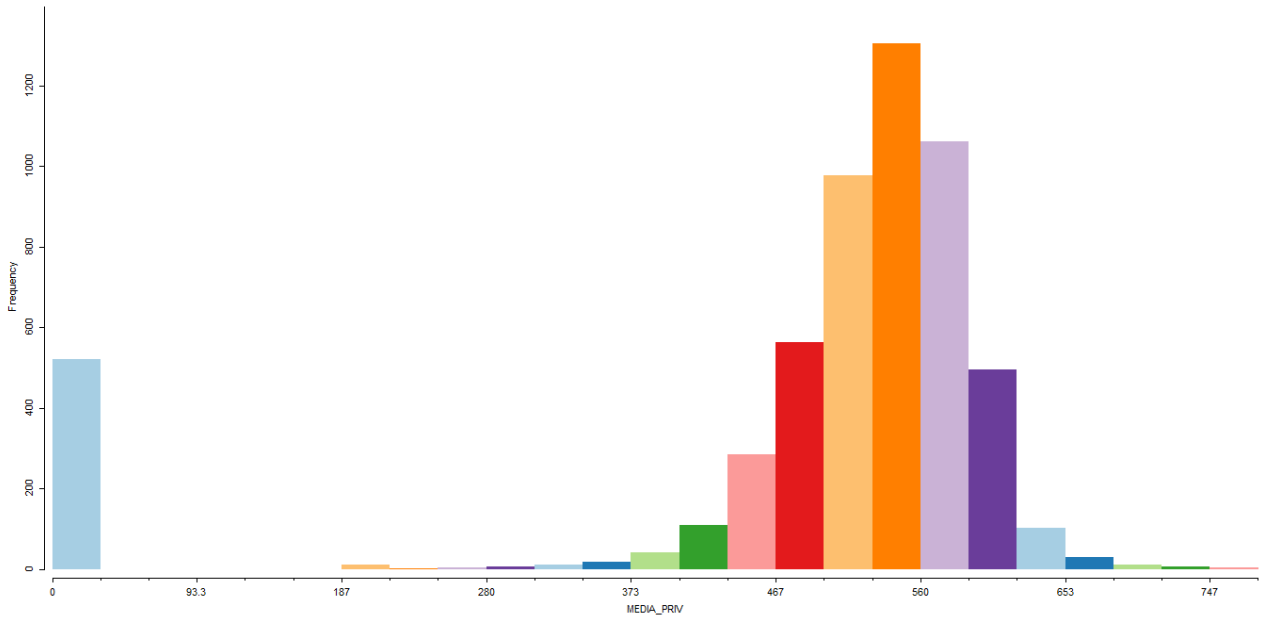


Imagem 19 – Histograma da média por estudo em Escola Privada



Foram analisadas também as médias com base no sexo dos inscritos. Inicialmente foi possível visualizar que os homens tem maior concentração nas médias, enquanto para as mulheres as médias tem uma maior amplitude, porém, ainda sem possibilidades de visualizar facilmente algum desempenho maior destacado.

Imagem 20 – Box-plot da média do sexo Feminino

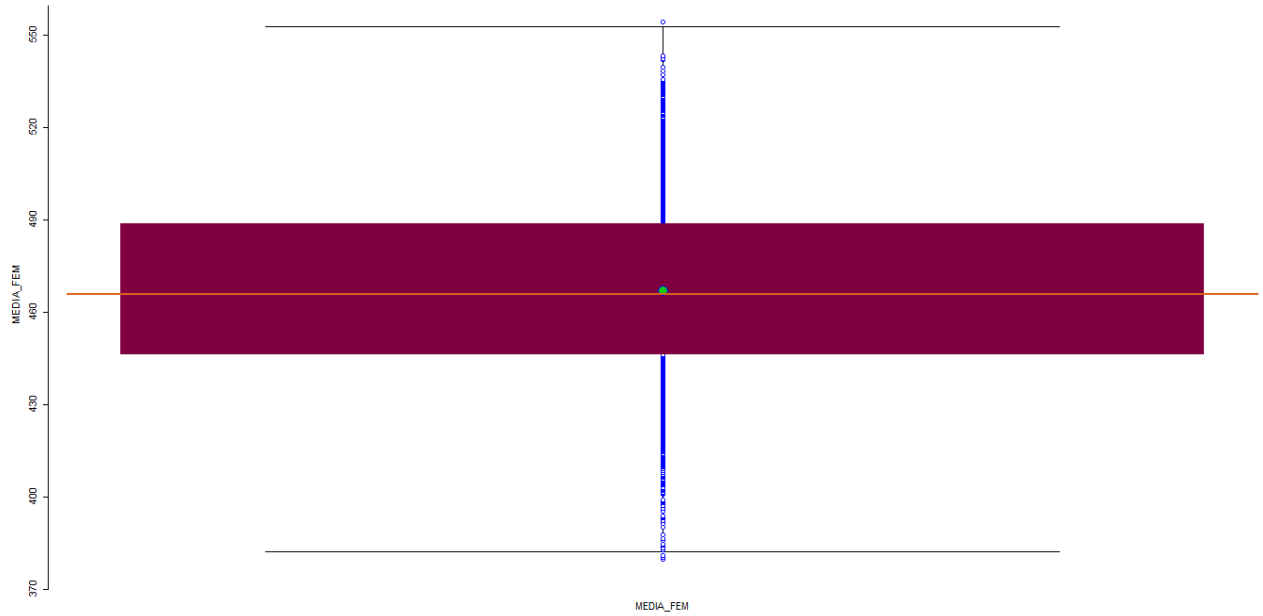


Imagem 21 – Histograma da média geral do sexo Feminino

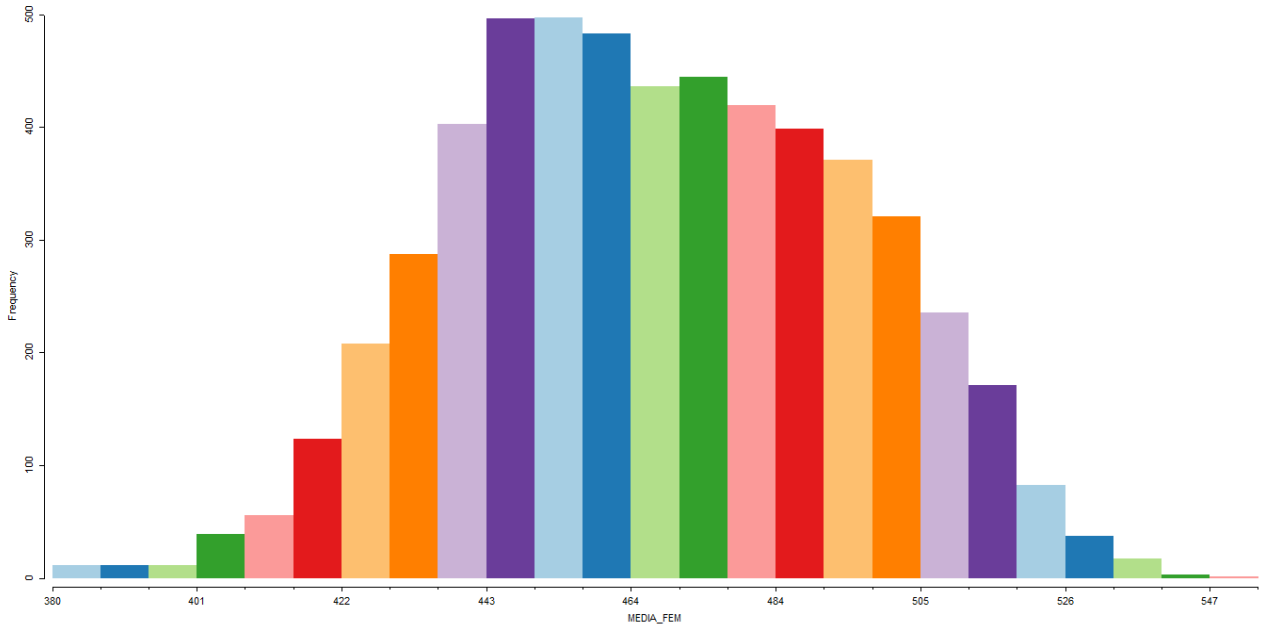


Imagem 22 – Box-plot da média do sexo Masculino

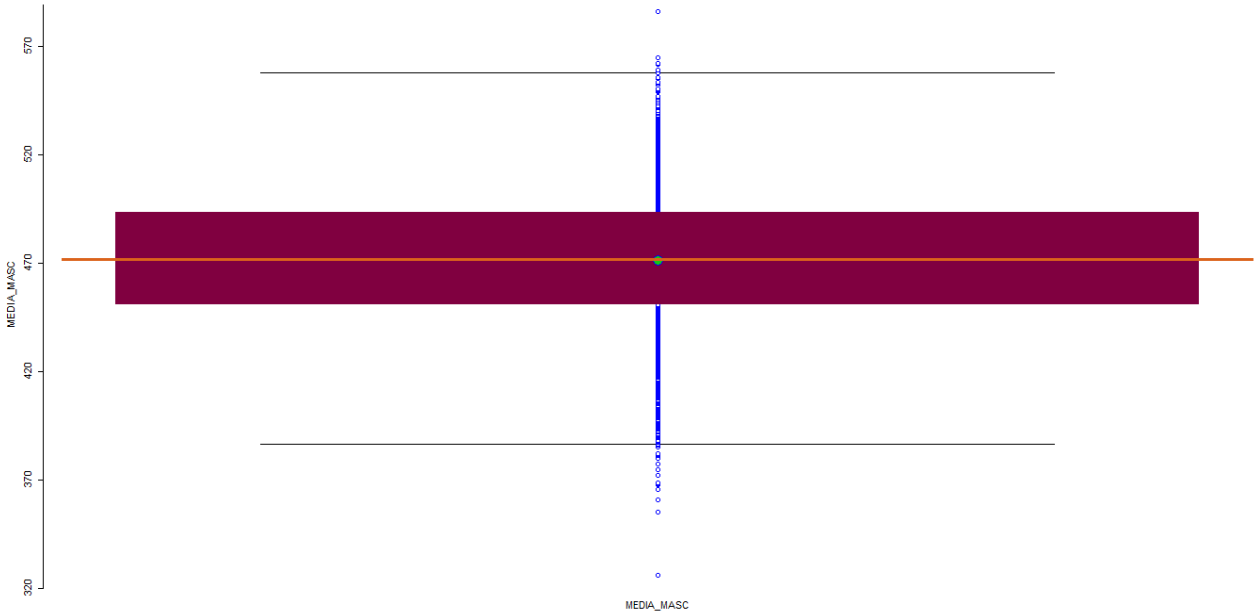
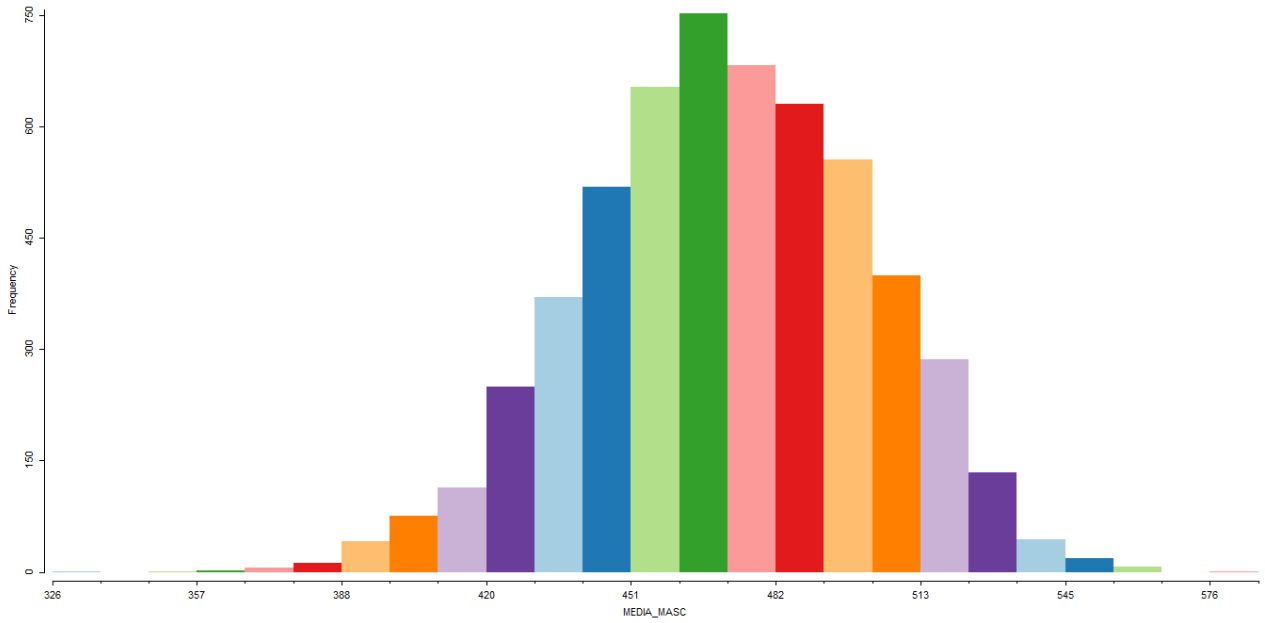
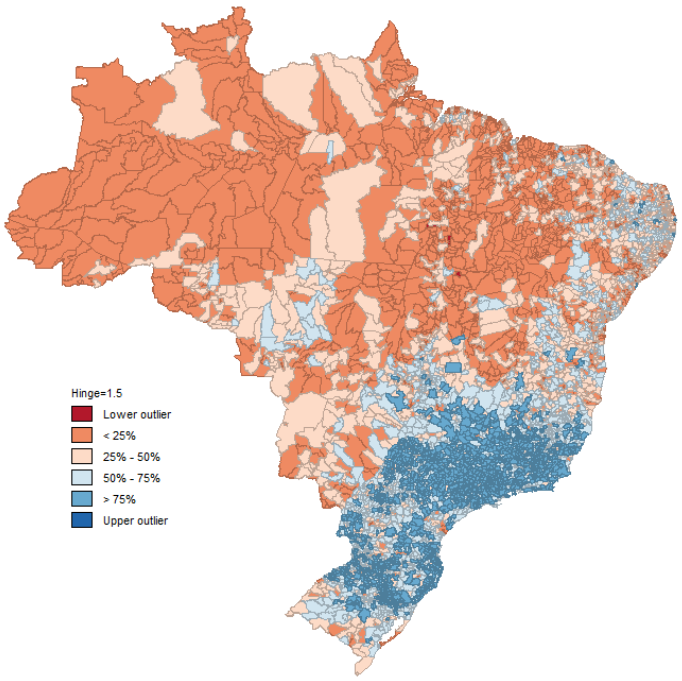


Imagem 23 – Histograma da média geral do sexo Masculino



Dadas as primeiras constatações obtidas nos box-plots e histogramas, foram analisadas as mesmas médias e variáveis de forma geográfica, buscando entender sua distribuição regional. Analisando a média geral nos municípios, visualiza-se em tons de azul os melhores desempenhos (municípios com resultados 75% acima da média), e em tons de vermelho os menores (municípios com resultados abaixo de 25% da média). É possível identificar facilmente um forte domínio das regiões Sudeste e Sul como altos desempenho, e regiões Norte e parte do Nordeste como baixos desempenho.

Imagem 24 – Box map da Média Geral



Analisando o cenário de estudantes ou concluintes em escolas públicas e privadas geograficamente, visualizamos em tons azuis os melhores desempenhos (resultados 75% acima da média), e vermelhos os menores (resultados abaixo de 25% da média), e identificamos como o desempenho dos que estudaram em escolas privadas tem melhor desempenho dada a predominância maior de azul na **Imagem 26**, que aparenta inclusive em maior quantidade nas regiões Nordeste e Norte. Também surgem na mesma imagem mais pontos em vermelho escuro, porém, estes não indicam piores desempenhos, na verdade tratam-se de municípios que não houve registros de notas de alunos de escolas privadas.

Imagem 25 – Box Map da Média por estudo em Escola Pública

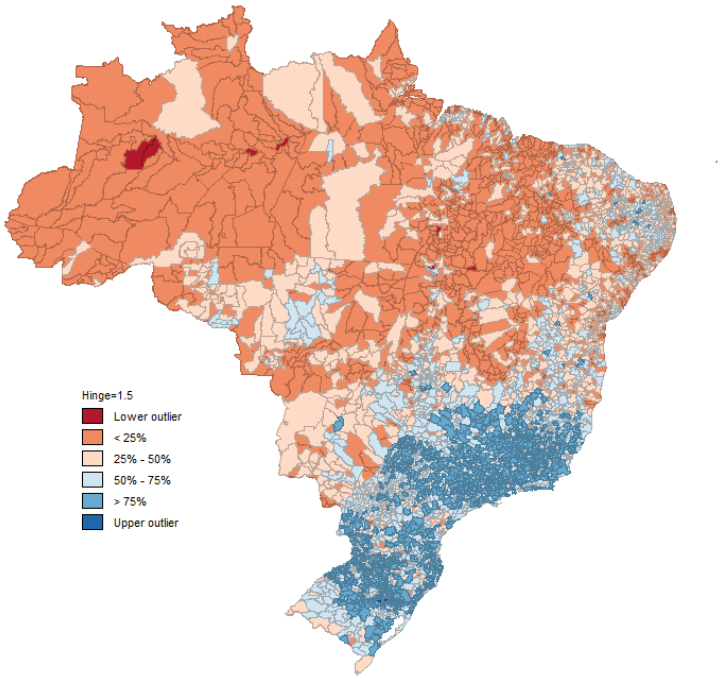
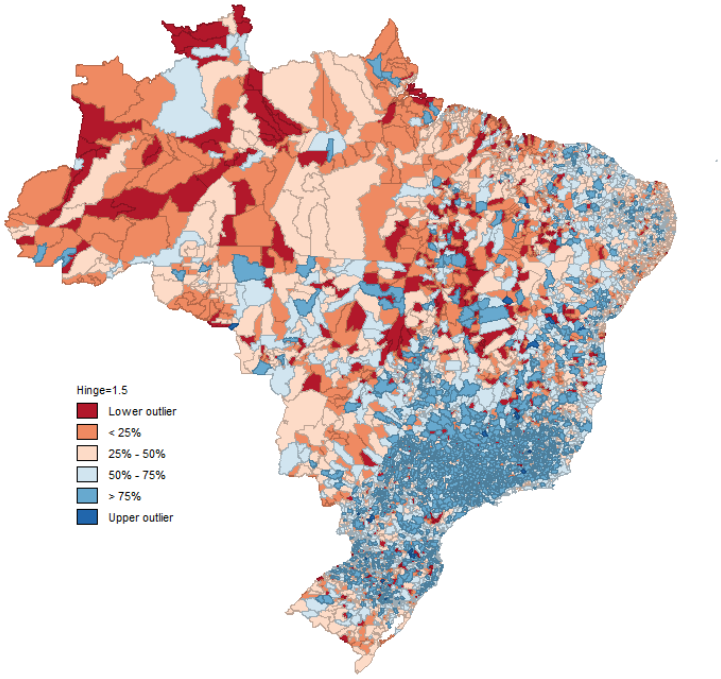


Imagem 26 – Box Map da Média por estudo em Escola Privada



Analisando o cenário de estudantes pelo sexo geograficamente, visualizamos que a diferença não é facilmente perceptível entre os sexos, mas em especial os homens apresentam em alguns municípios desempenho 25% abaixo da média, conforme tons vermelhos escuros na **Imagem 28**, enquanto as mulheres mantem uma média mais normal, sem discrepâncias aparentes.

Imagem 27 – Box Map da Média para sexo Feminino

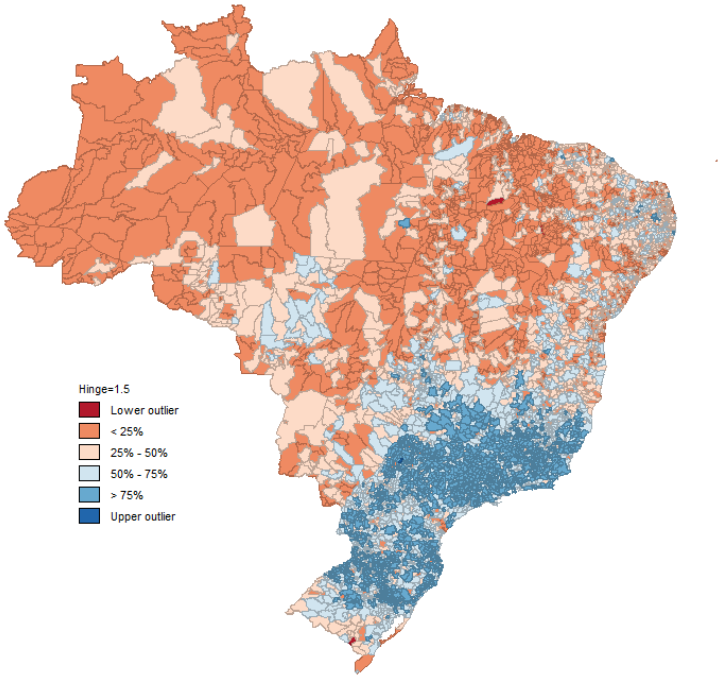
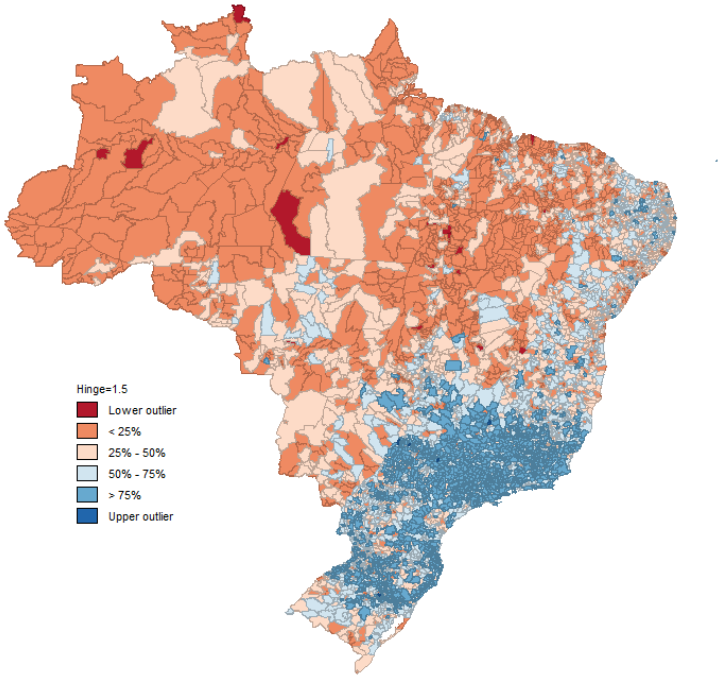


Imagem 28 – Box Map da Média para sexo Masculino



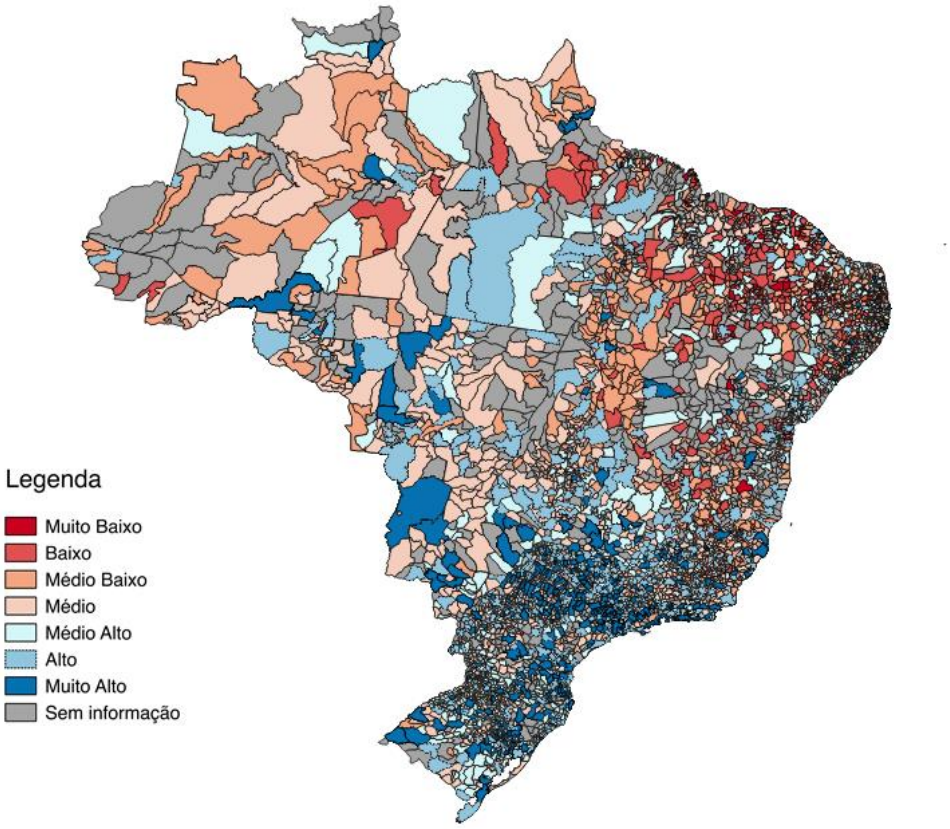
4. DEFINIÇÃO DE HIPÓTESES

4.A – VISÃO POR ESCOLAS

Com base nas analises feitas na sessão 3.A, foi possível notar que as notas médias das escolas aparentam estar relacionadas com algo além dos municípios. Observando os mapas e os histogramas divididos por dependência administrativa, principalmente analisando instituições publicas x privadas, foi possível levantar a seguinte hipótese: Será que situação socioeconômica dos alunos afeta as notas do Enem em suas escolas? Ou seja, se a maioria predominante de determinada escola for composta por alunos de classe mais alta, é provável que a nota média daquela instituição seja melhor do que a nota média de escolas onde os alunos, em sua maioria, sejam de classe mais baixa?

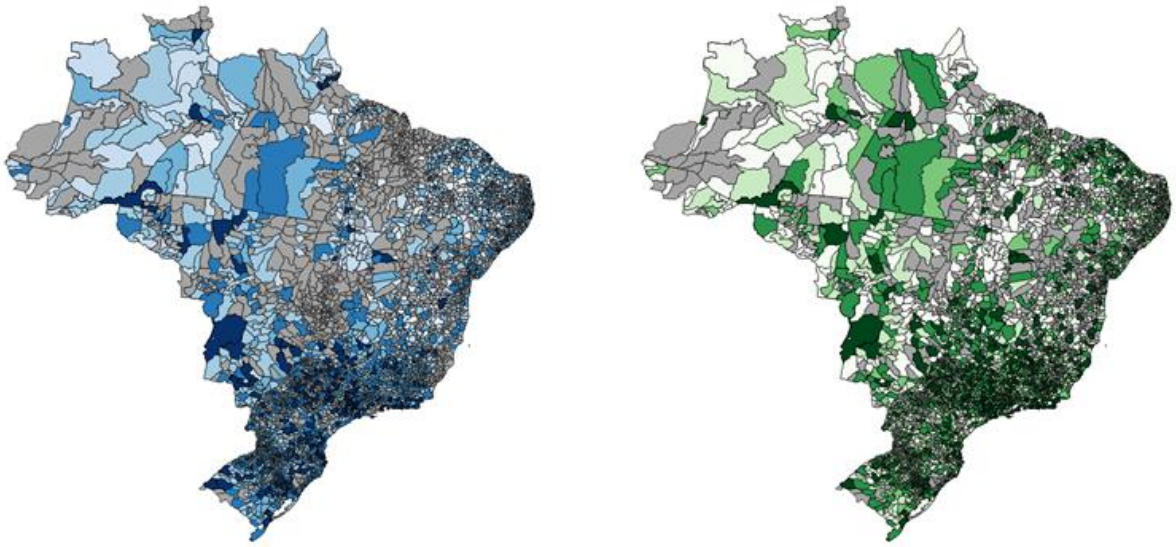
Para validar essa hipótese, utilizou-se a variável Inse (indicador de nível socioeconômico) encontrada na planilha de escolas. As técnicas e cálculos dessa variável pode ser encontradas com mais detalhes em:<http://download.inep.gov.br/educacao_basica/enem/enem_por_escola/2015/nota_tecnica_indicador_nivel_socioeconomico.pdf>.
Abaixo, observa-se um mapa com a classificação socioeconômica por municípios.

Imagem 29 – Situação socioeconômica por municípios



Comparando o mapa de notas por munícipios, com o mapa da situação socioeconômica por municípios (Figura 30), é possível observar uma grande semelhança, apontando uma possível correlação entre estas duas variáveis.

Imagem 30 – Comparação notas x situação socioeconômica



4.B – VISÃO POR INDIVÍDUOS

Após as análises apresentadas na seção 3.B, visualizamos as primeiras evidências:

1. As regiões Sul e Sudeste apresentam maior número de municípios com desempenho acima de 75% da média, e alguns municípios em especial dos estados do Ceará, Pernambuco e Paraíba.
2. Dos indivíduos que realizara o ENEM, o desempenho daqueles que estudam em escolas privadas prevalece com melhor desempenho e mais distribuído regionalmente quando comparando com indivíduos que estudam em escolas públicas.
3. O desempenho das mulheres aparenta levemente melhor em relação ao desempenho dos homens, explicado por uma normalidade mais concentrada das médias, em relação à distribuição das médias dos homens, mas não a ponto de ser possível destacar qual dos grupos tem maior desempenho.

Mediante estas constatações, fez-se necessário avaliar a relação das variáveis de cada cenário e avaliar o quanto estes realmente influenciam na média geral do ENEM, e de que forma. Sendo assim, os seguintes gráficos de dispersão foram analisados para entender cada caso.

Imagem 31 – Dispersão da Média em Ensino Privado vs. Média Geral.

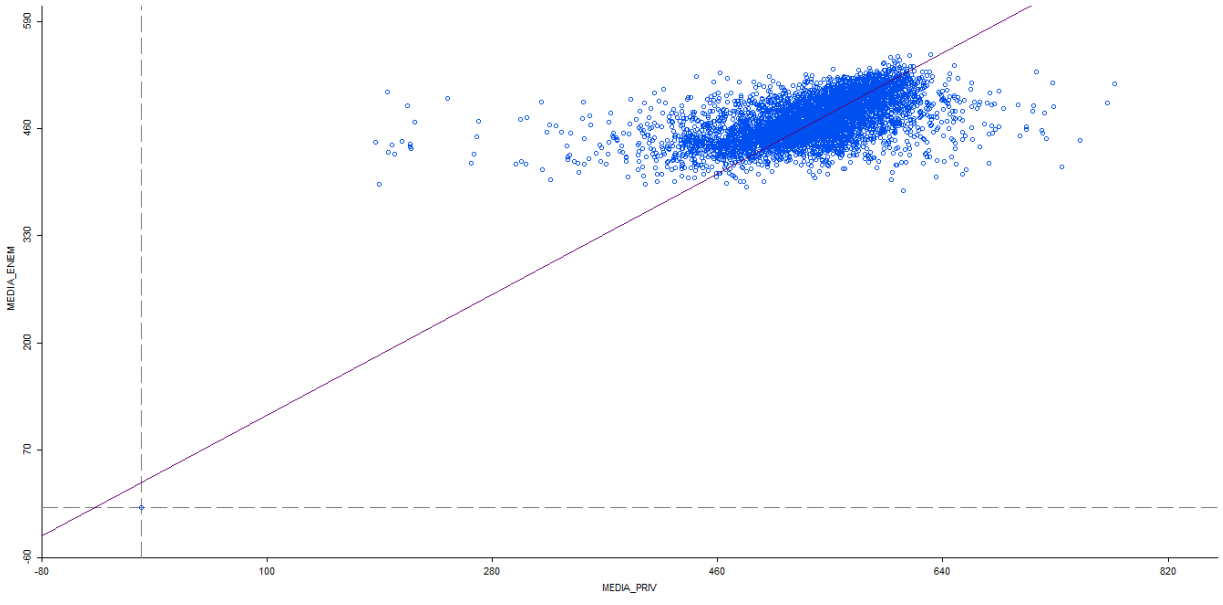


Imagem 32 – Dispersão da Média em Ensino Público vs. Média Geral.

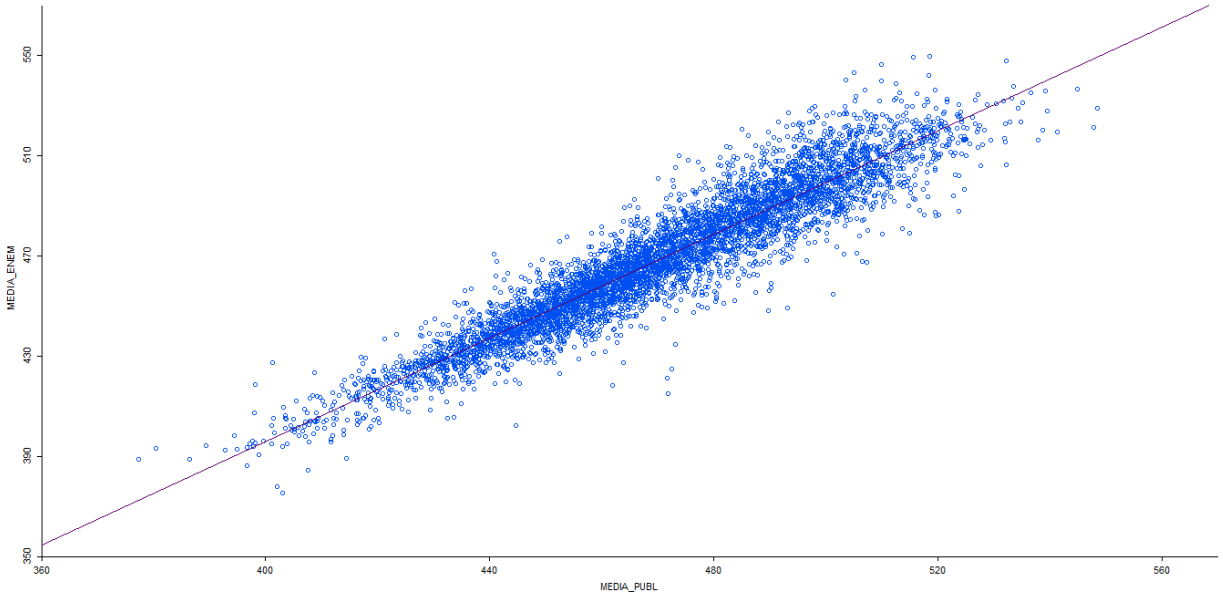


Imagem 33 – Dispersão da Média das Mulheres vs. Média Geral.

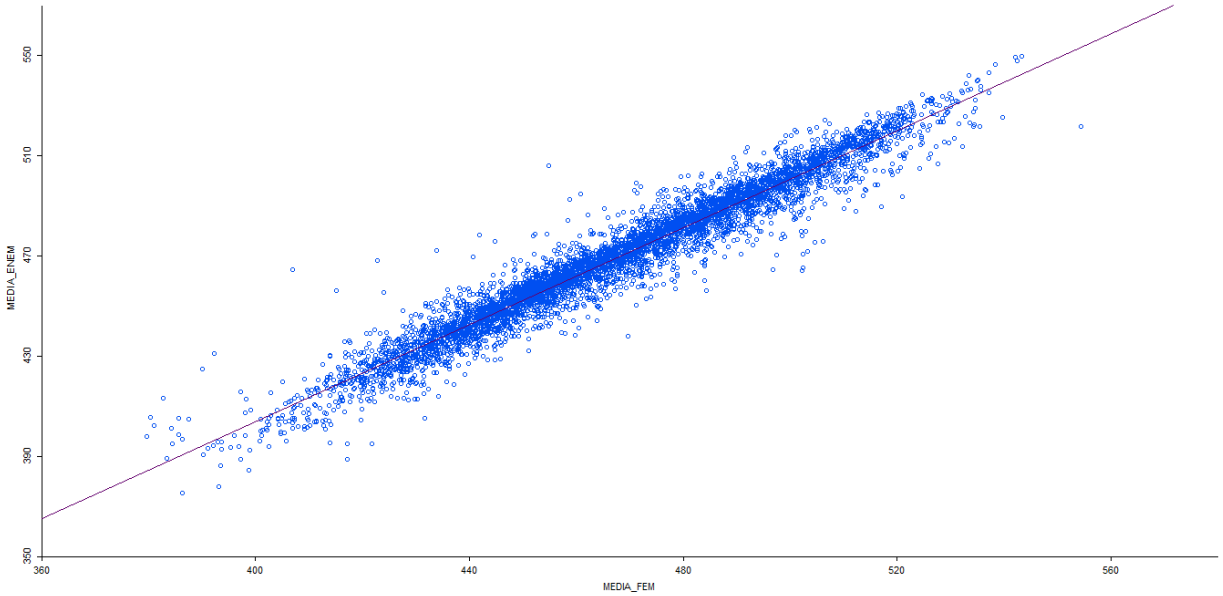
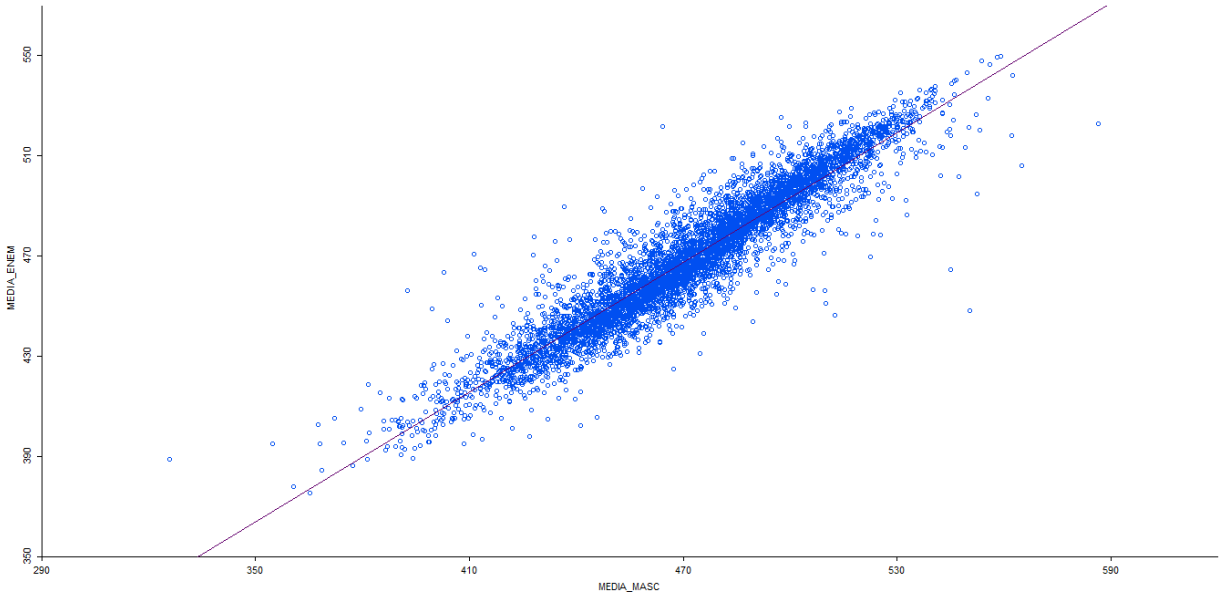


Imagem 34 – Dispersão da Média dos Homens vs. Média Geral.



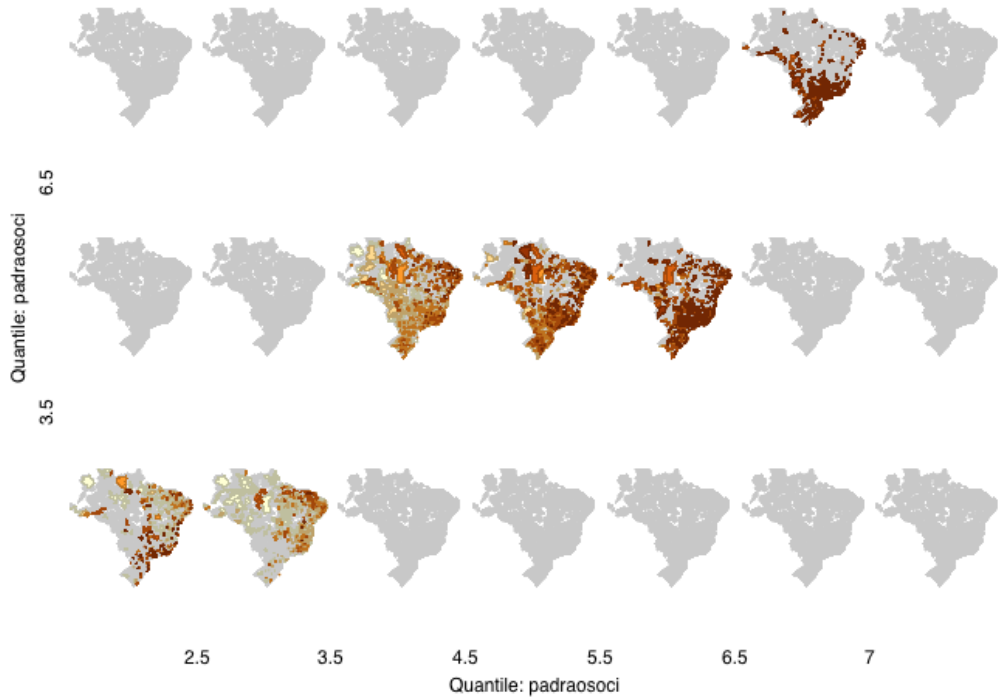
Após visualização individual das variáveis em relação à média geral, surgiu uma primeira evidência de que homens podem apresentar melhor desempenho ou reflexo na média geral em relação à mesma análise feita com a média das mulheres.

5. VALIDAÇÃO DE HIPÓTESES

5.A – VISÃO POR ESCOLAS

O primeiro passo para a validação da hipótese de que a situação socioeconômica dos alunos das escolas analisadas afeta diretamente suas médias gerais do Enem, foi montar um mapa para acompanhar esta evolução, como mostrado abaixo na Imagem 35.

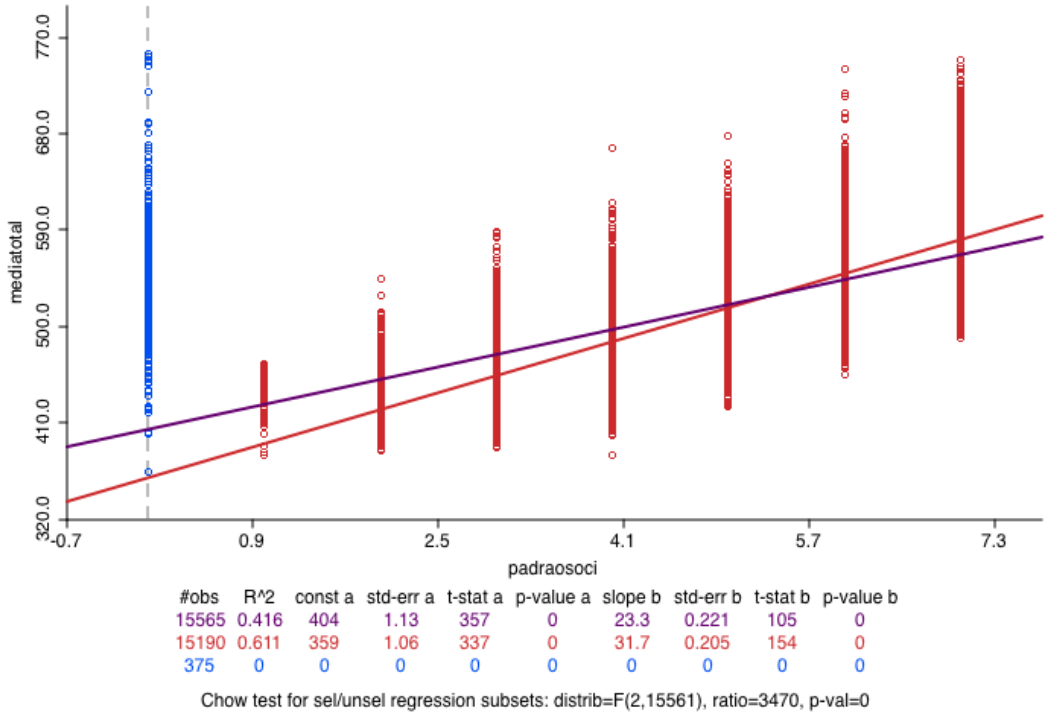
Imagem 35 – Evolução do nível socioeconômico e notas



No mapa acima fica evidente que, a medida que o padrão socioeconômico aumenta, as médias gerais das notas também aumentam.

A regressão apresentada na Figura 36 confirma esta mesma hipótese. Apesar de haver exceções, é possível observar com clareza uma tendência de aumento nas notas conforme a classificação socioeconômica também aumenta. Uma boa parte da variação nas notas pode ser explicada pela variação do padrão socioeconômico.

Imagem 36 – Regressão padrão socioeconômico x notas



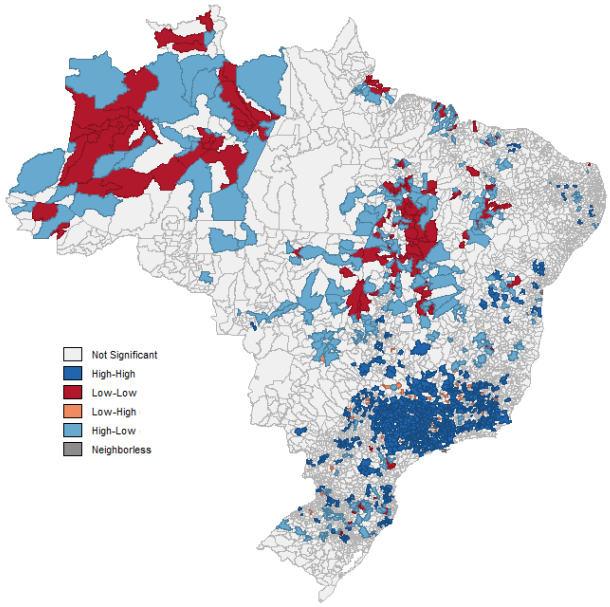
5.B – VISÃO POR INDIVÍDUOS

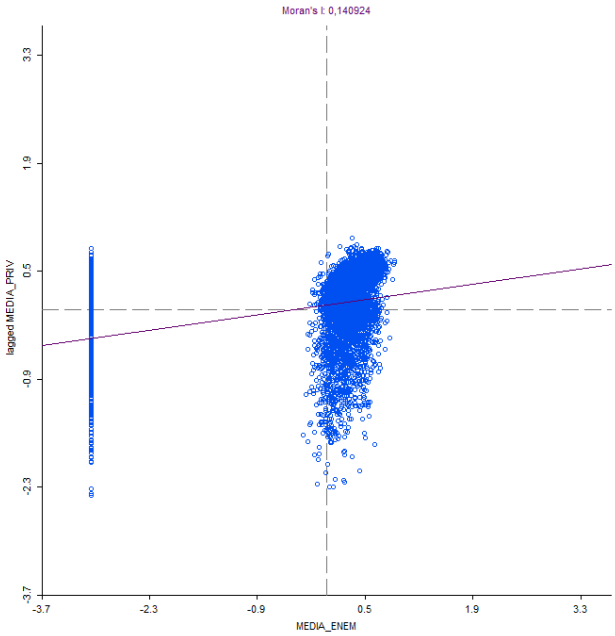
Para entender de maneira mais quantitativa o quanto cada hipótese influencia diretamente na média geral do ENEM, realizamos a técnica de regressão linear modelando a média geral por cada variável individualmente, e demonstramos os resultados conforme a seguir:

a-) Média dos indivíduos de Escola Privada e sua inferência na média geral do ENEM, juntamente com análise de cluster.

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  munic_priv
Dependent Variable  :  MEDIA_ENEM  Number of Observations: 5560
Mean dependent var  :     426,608  Number of Variables   :    2
S.D. dependent var  :      139,74  Degrees of Freedom    : 5558

R-squared           :    0,923178  F-statistic           :     66791,4
Adjusted R-squared  :    0,923164  Prob(F-statistic)     :           0
Sum squared residual:8,34067e+006  Log likelihood        :    -28220,3
Sigma-square        :     1500,66  Akaike info criterion :     56444,5
S.E. of regression  :     38,7384  Schwarz criterion     :     56457,8
Sigma-square ML     :     1500,12
S.E of regression ML:     38,7314

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------
```

```
    CONSTANT       31,67044       1,614053       19,62169     0,0000000
  MEDIA_PRIV     0,8136505    0,003148312       258,4402     0,0000000
------------------------------------------------------------------------


REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   6,048272
TEST ON NORMALITY OF ERRORS
TEST                   DF          VALUE            PROB
Jarque-Bera             2          10156,58         0,0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF          VALUE            PROB
Breusch-Pagan test      1          154,7595         0,0000000
Koenker-Bassett test    1          36,73904         0,0000000
SPECIFICATION ROBUST TEST
TEST                   DF          VALUE            PROB
White                   2          52,78895         0,0000000
========================= END OF REPORT ==============================
```

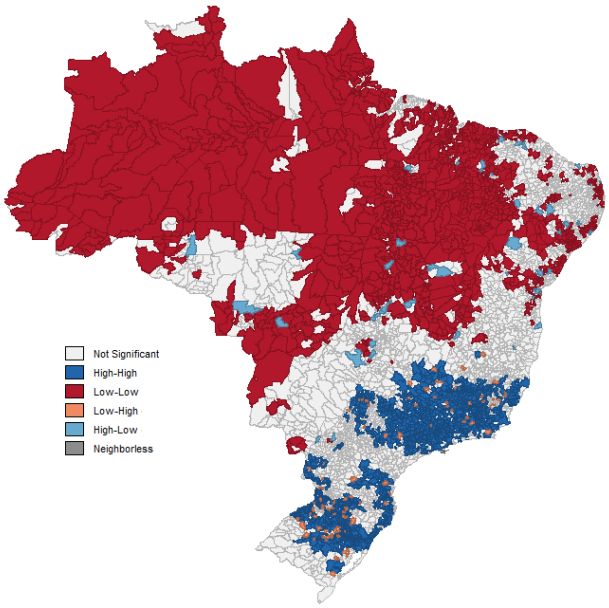


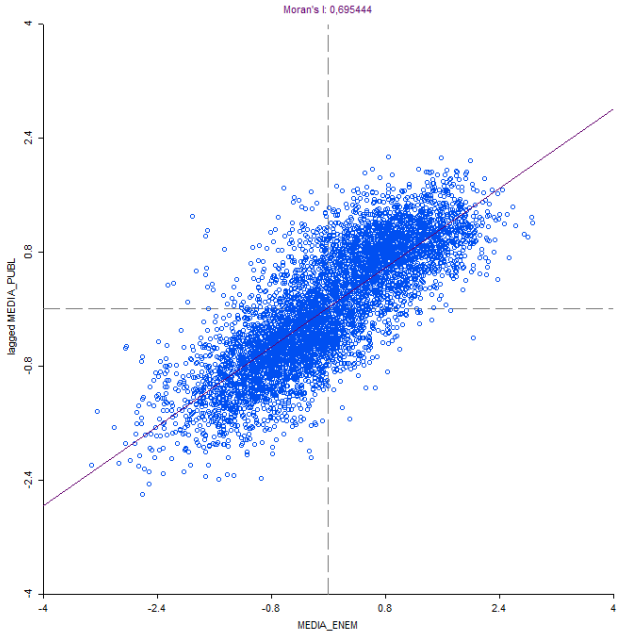


b-) Média dos indivíduos de Escola Pública e sua inferência na Média Geral do ENEM.

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  munic_publ
Dependent Variable  :  MEDIA_ENEM  Number of Observations: 5560
Mean dependent var  :     468,768  Number of Variables   :    2
S.D. dependent var  :     28,1536  Degrees of Freedom    : 5558

R-squared           :    0,885766  F-statistic           :     43096,4
Adjusted R-squared  :    0,885745  Prob(F-statistic)     :           0
Sum squared residual:      503429  Log likelihood        :    -20415,5
Sigma-square        :     90,5773  Akaike info criterion :     40835,1
S.E. of regression  :     9,51721  Schwarz criterion     :     40848,3
Sigma-square ML     :     90,5447
S.E of regression ML:      9,5155

------------------------------------------------------------------------
```

```
    Variable   Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------
    CONSTANT     -17,37748       2,345253      -7,409639    0,0000000
  MEDIA_PUBL      1,033172    0,004976824       207,5968    0,0000000
------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   36,721930
TEST ON NORMALITY OF ERRORS
TEST                   DF          VALUE            PROB
Jarque-Bera             2           669,4461        0,0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF          VALUE            PROB
Breusch-Pagan test      1           302,8736        0,0000000
Koenker-Bassett test    1           163,8246        0,0000000
SPECIFICATION ROBUST TEST
TEST                   DF          VALUE            PROB
White                   2           170,9686        0,0000000
========================= END OF REPORT ==============================
```

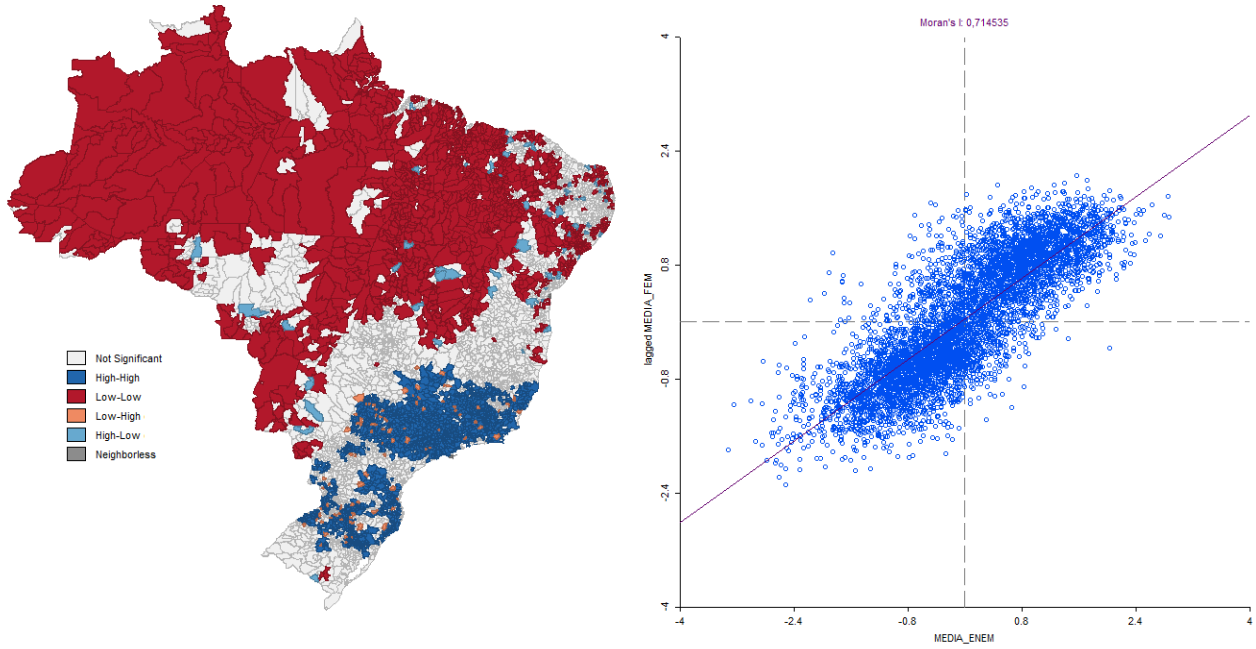




c-) Média das Mulheres e sua inferência na Média Geral do ENEM.

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  munic_enem
Dependent Variable  :  MEDIA_ENEM  Number of Observations: 5560
Mean dependent var  :     468,768  Number of Variables   :    2
S.D. dependent var  :     28,1536  Degrees of Freedom    : 5558

R-squared           :    0,943815  F-statistic           :     93364,6
Adjusted R-squared  :    0,943805  Prob(F-statistic)     :           0
Sum squared residual:      247608  Log likelihood        :    -18442,9
Sigma-square        :     44,5498  Akaike info criterion :     36889,7
S.E. of regression  :     6,67457  Schwarz criterion     :       36903
Sigma-square ML     :     44,5338
S.E of regression ML:     6,67337

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
```

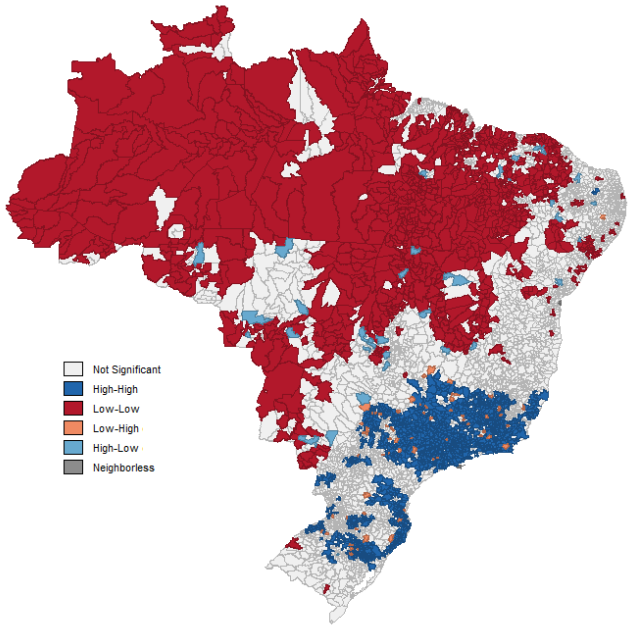
```
------------------------------------------------------------------------
    CONSTANT       16,66856          1,4823        11,24507    0,0000000
   MEDIA_FEM      0,9678933     0,003167644        305,5562    0,0000000
------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   33,089024
TEST ON NORMALITY OF ERRORS
TEST                     DF          VALUE            PROB
Jarque-Bera               2           4487,118        0,0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                     DF          VALUE            PROB
Breusch-Pagan test        1           2,866725        0,0904287
Koenker-Bassett test      1          0,8976183        0,3434211
SPECIFICATION ROBUST TEST
TEST                     DF          VALUE            PROB
White                     2           72,96882        0,0000000
========================= END OF REPORT ==============================
```

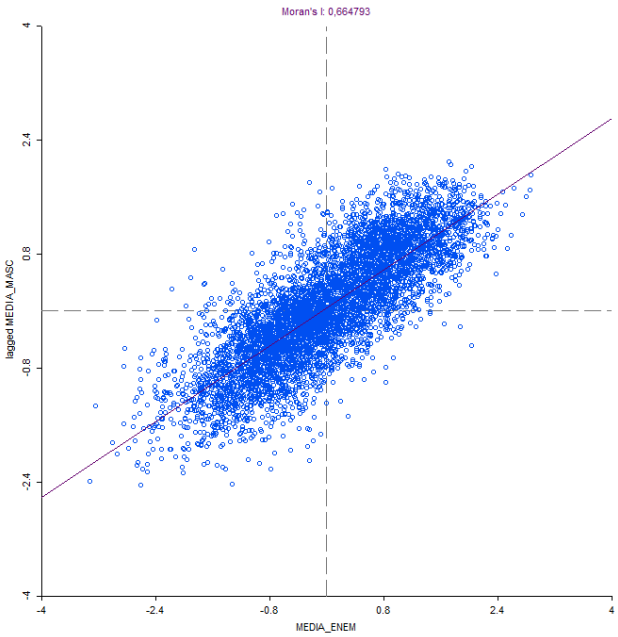


d-) Média dos Homens e sua inferência na Média Geral do ENEM.

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  munic_enem
Dependent Variable  :  MEDIA_ENEM  Number of Observations: 5560
Mean dependent var  :     468,768  Number of Variables   :    2
S.D. dependent var  :     28,1536  Degrees of Freedom    : 5558

R-squared           :    0,885305  F-statistic           :       42901
Adjusted R-squared  :    0,885284  Prob(F-statistic)     :           0
Sum squared residual:      505459  Log likelihood        :    -20426,7
Sigma-square        :     90,9426  Akaike info criterion :     40857,5
S.E. of regression  :     9,53638  Schwarz criterion     :     40870,7
Sigma-square ML     :     90,9099
S.E of regression ML:     9,53467

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
```

```
------------------------------------------------------------------------
    CONSTANT       62,14652       1,967326       31,58933    0,0000000
  MEDIA_MASC      0,8623401    0,004163371       207,1255    0,0000000
------------------------------------------------------------------------


REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   30,732679
                                      (Extreme Multicollinearity)
TEST ON NORMALITY OF ERRORS
TEST                     DF          VALUE            PROB
Jarque-Bera               2           5207,901        0,0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                     DF          VALUE            PROB
Breusch-Pagan test        1         0,04973728        0,8235211
Koenker-Bassett test      1          0,0147587        0,9033066
SPECIFICATION ROBUST TEST
TEST                     DF          VALUE            PROB
White                     2           101,3267        0,0000000
========================= END OF REPORT ==============================
```





Com as regressões apresentadas, conseguimos constatar as hipóteses, porém, como tratam-se de regressões aplicadas com os dados das análises geográficas onde as médias estão computadas nas variáveis agrupadas pelos municípios, foi detectada multicolinearidade conforme os testes e destaques da regressão. Devido a este fator, uma nova base foi organizada baseada nos inscritos e suas médias individuais, sem considerar os municípios, e foram geradas variáveis dummies para distinguir escolas públicas de privadas, sexos feminino de masculino, e assim aplicada uma regressão linear múltipla para se obter melhores resultados, além total controle do modelo sem o fator da multicolinearidade. Os resultados desta nova regressão são apresentados a seguir:

```
Call:
lm(formula = MEDIA ~ PRIVADA + FEM, data = Dados)

Residuals:
    Min      1Q  Median      3Q     Max
-520.57  -47.01    6.33   59.21  360.47
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 479.45452    0.06042  7935.4   <2e-16 ***
PRIVADA      85.11337    0.11292   753.8   <2e-16 ***
FEM          -9.10112    0.07672  -118.6   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 94.26 on 6192658 degrees of freedom
Multiple R-squared:  0.08672,  Adjusted R-squared:  0.08672
F-statistic: 2.94e+05 on 2 and 6192658 DF,  p-value: < 2.2e-16
```

Com estes resultados, pode-se facilmente simular os cenários e assim temos como exemplos os casos abaixo:

```
de Escola Privada e Sexo Feminino  = 555,466772
de Escola Privada e Sexo Masculino = 564,567888
de Escola Pública e Sexo Feminino  = 470,353401
de Escola Pública e Sexo Masculino = 479,454517
```

6. CONCLUSÕES

Sabendo da importância do Enem como medida de desempenho do ensino básico, em especial o ensino médio, o trabalho se propôs a entender como os dados se comportam, para explicar através de análises, levantamento e validações de hipóteses, quais variáveis influenciam a média geral dos inscritos e das instituições de ensino.

Apesar de duas visões distintas (por escolas e por indivíduos) é possível notar que as hipóteses e as conclusões a que chegamos são as mesmas. Levando-se em consideração todos os aspectos expostos através de gráficos, mapas e regressões, somos levados a crer que:

- As notas sofrem influência do meio, ou seja, existe uma autocorrelação espacial onde as notas mostram padrões de acordo com o município.

- O desempenho de alunos de escolas privadas mostra-se superior ao desempenho dos alunos de escolas públicas. Padrão que se repete na visão da média do desempenho das escolas. A situação socioeconômica dos alunos de cada escola mostra-se como forte fator de influência na média geral.

Como próximos passos para enriquecer, complementar e/ou contrapor o presente estudo, sugerimos uma analise com outros indicadores socioeconômicos como PIB, IDH, Coeficiente de Gini e taxa de desemprego.