

Pós-graduação em Ciência de Dados

***Disciplina*** – Métodos Estatísticos de Apoio a Decisão I

**(*Laboratório 01 – SAS Guide e Introdução a Análise Estatística*)**

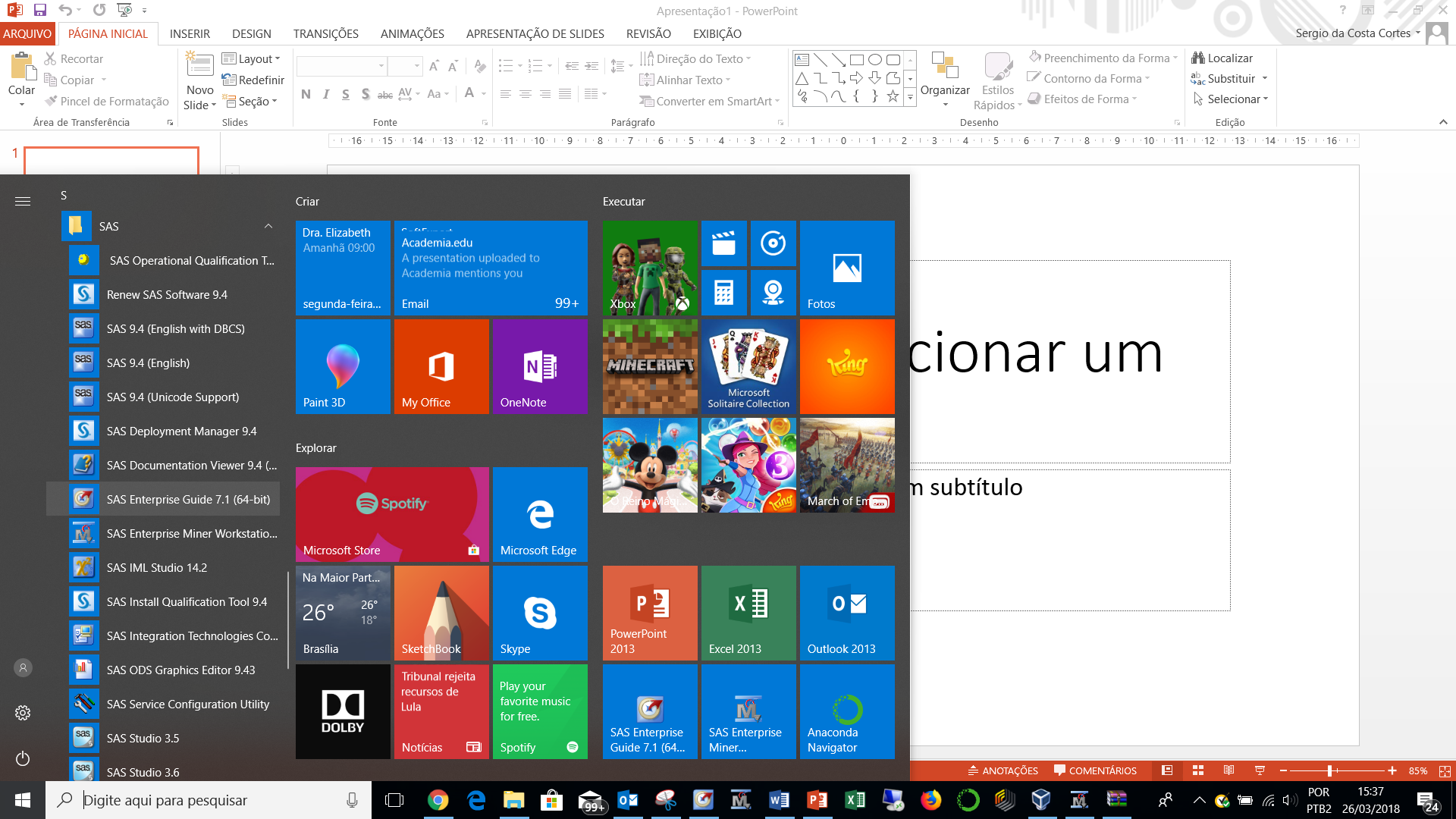
***Objetivos do Laboratório*** – Familiarizar o aluno com a utilização do software *SAS Guide* e com os princípios básicos da Análise Estatística

**Prof. Sérgio Côrtes**

**Prof. Suélio Moura**

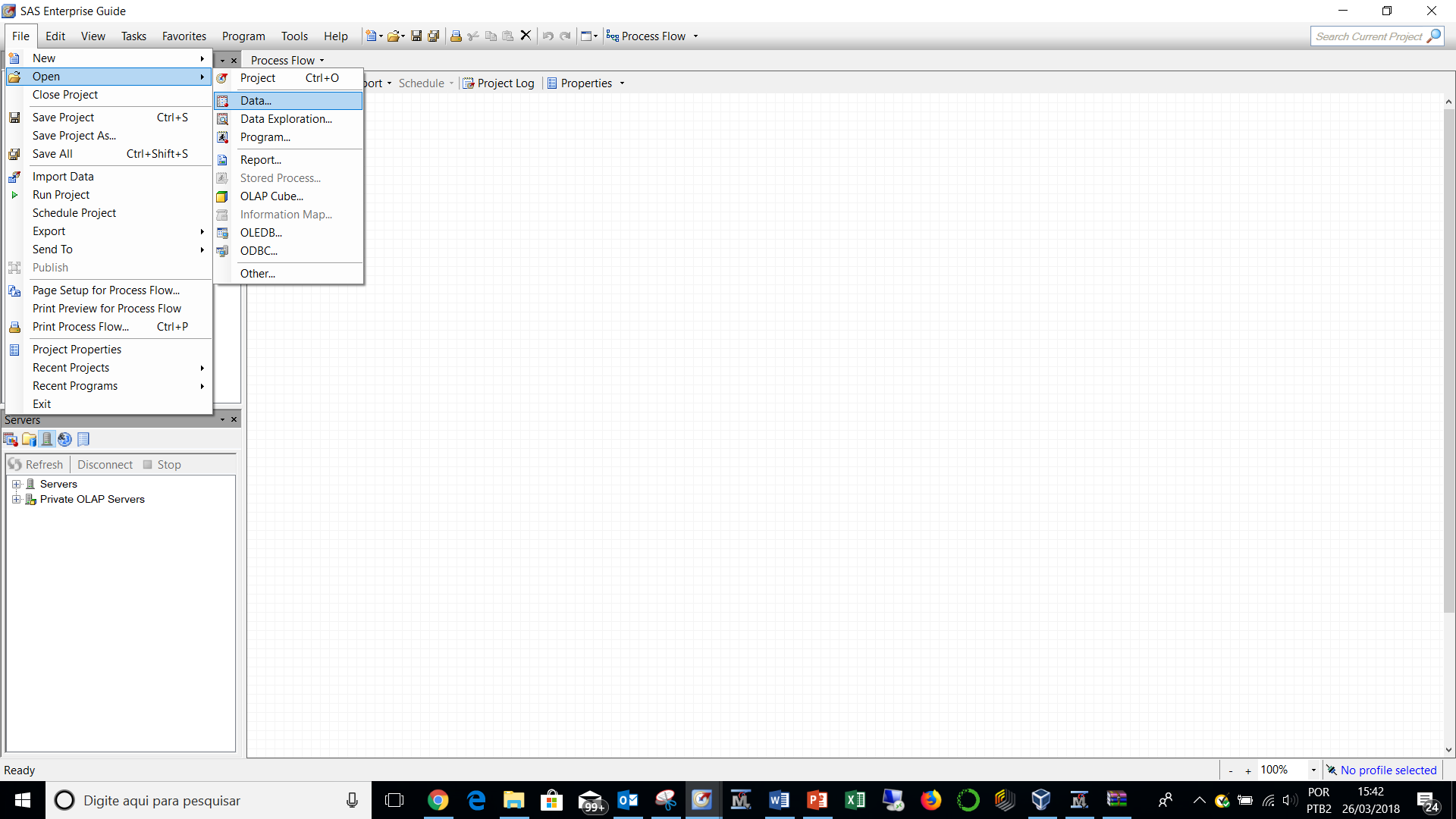
**1a versão**

**Agosto de 2018**

1. No menu de programas do Windows Inicialize o programa SAS Guide 
2. Selecione a opção New Project



1. Selecione a opção (File🡪 Open 🡪 Data) e selecione o *Data Set SAS* “*dados\_enem\_amostra\_aula\_selec”* na pasta “CIA:\DADOS\ENEM” ou onde você armazena os arquivos de trabalho da disciplina.



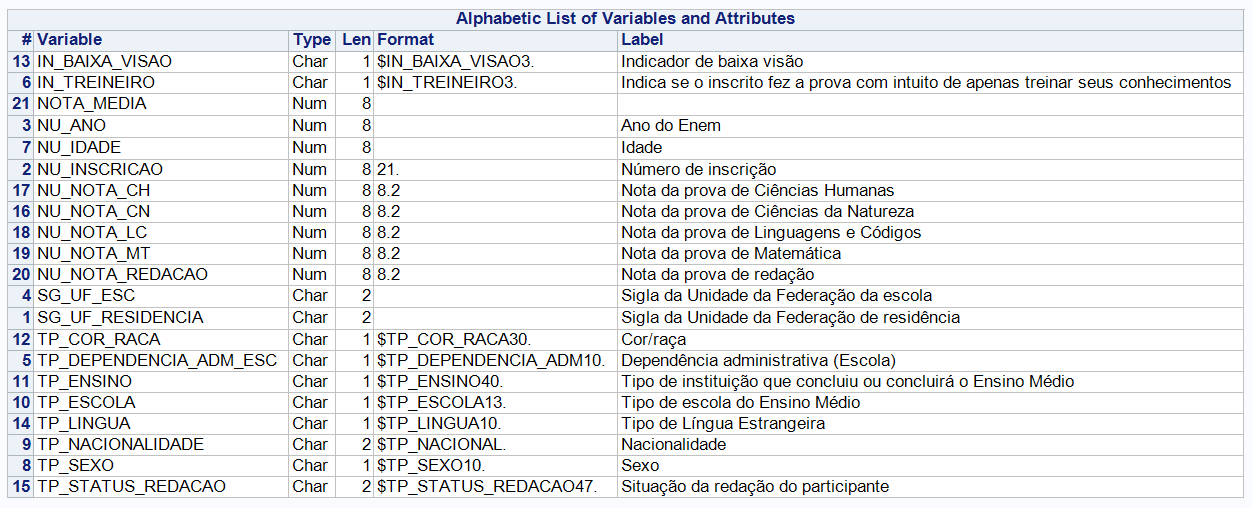
1. Execute o comando “Proc Contents” e verifique as variáveis deste arquivo:

File 🡪 New 🡪 Program

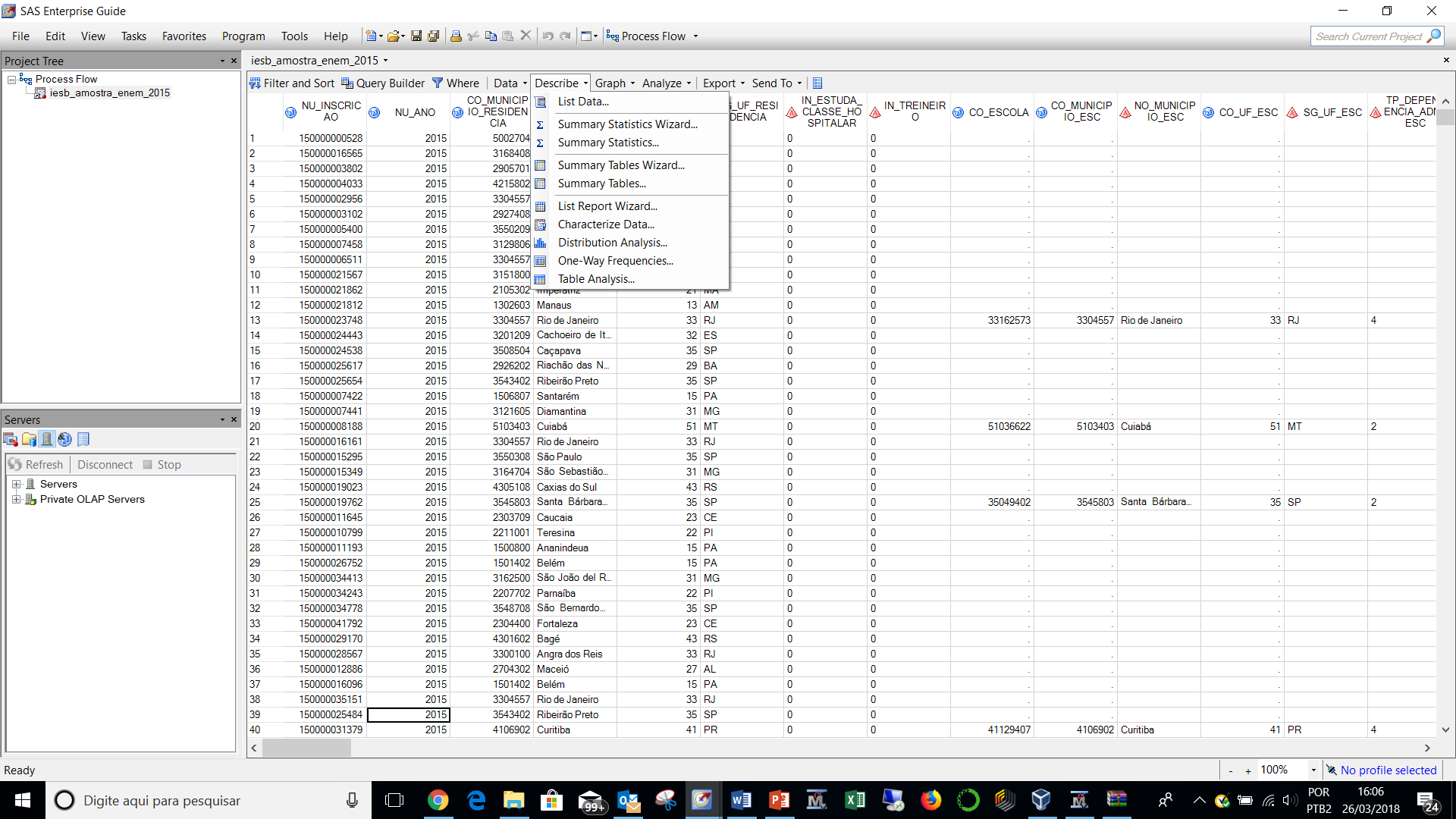
**proc** **contents** data="CIA:\Dados\ENEM\dados\_enem\_amostra\_aula\_selec";

Run;

Resultado:



1. Utilizando o *Dataset SAS* “*dados\_enem\_amostra\_aula\_selec*”, vamos utilizar as tarefas do componente *Describe* do *SAS Guide* para conhecermos os dados. Conheça todas as opções deste menu.



1. Selecione o componente One-Way Frequencies.
   * 1. Faça uma análise das variáveis Sexo (*TP\_Sexo)*, Cor/Raça (*TP\_COR\_RACA*) e Situação da Redação do Participante (*TP\_Status\_Redacao*). Selecione como estatísticas apenas *freqüências*. Escolha um modelo de gráfico para representar as variáveis*. O que você pode afirmar da análise destas variáveis? Avalie estas variáveis também por unidade da federação.*

**TP\_SEXO**

M ='Masculino'

F ='Feminino';

**TP\_COR\_RACA**

**0**='Não declarado'

**1**='Branca'

**2**='Preta'

**3**='Parda'

**4**='Amarela'

5='Indígena';

**TP\_STATUS\_REDACAO**

**1**='Sem problemas'

**2**='Anulada'

**3**='Cópia Texto Motivador'

**4**='Em Branco'

**6**='Fuga ao tema'

**7**='Não atendimento ao tipo'

**8**='Texto insuficiente'

9='Parte desconectada';

TP\_Ensino

|  |  |
| --- | --- |
| 1 | Ensino Regular |
| 2 | Educação Especial - Modalidade Substitutiva |
| 3 | Educação de Jovens e Adultos |

TP\_NACIONALIDADE

|  |  |
| --- | --- |
| 0 | Não informado |
| 1 | Brasileiro(a) |
| 2 | Brasileiro(a) Naturalizado(a) |
| 3 | Estrangeiro(a) |
| 4 | Brasileiro(a) Nato(a), nascido(a) no exterior |

TP\_ESCOLA

|  |  |
| --- | --- |
| 1 | Não Respondeu |
| 2 | Pública |
| 3 | Privada |
| 4 | Exterior |

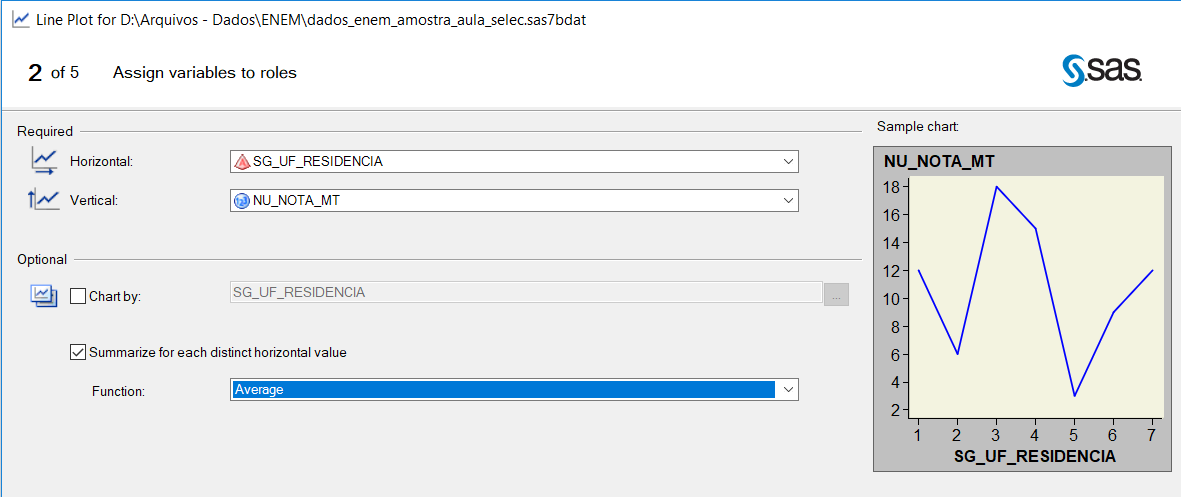
Resultados: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Utilizando o resultado da “Proc Contents” (item 4), escolha outro conjunto de variáveis para estudo. Explore bem essa opção, mude as frequências produzidas, personalize os títulos dos seus relatórios. *Avalie estas variáveis também por unidade da federação.*

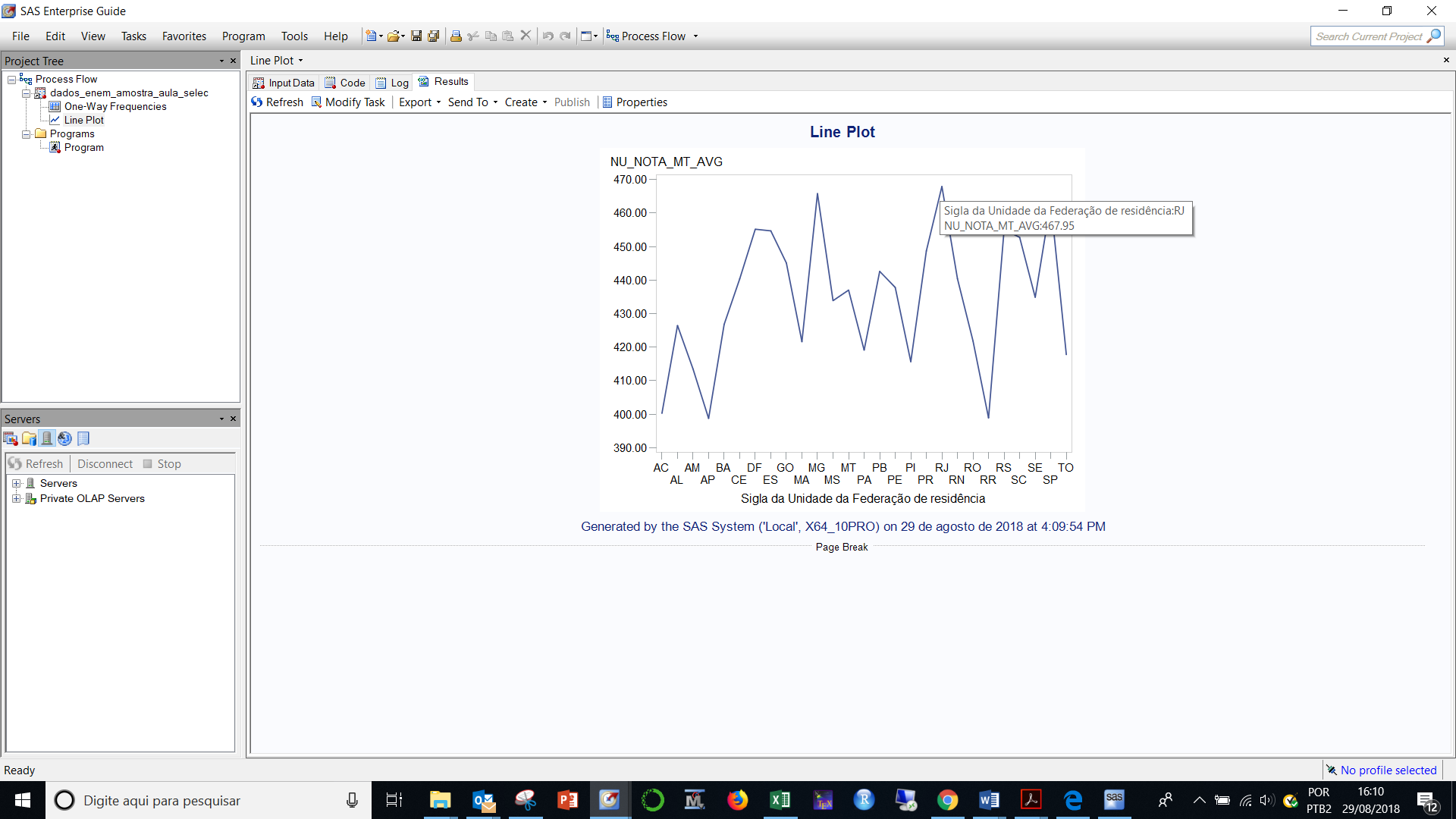
Resultados: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Selecione o componente Graph.
   1. Vamos construir um gráfico de linhas utilizando a opção “Line Plot Wizard”;
   2. Construa um gráfico de linhas para identificar as notas médias da nota de Matemática (NU\_NOTA\_MT) por Unidade da Federação de Residência do Candidato (SG\_UF\_Residencia);

Sugestão:



Resultado:

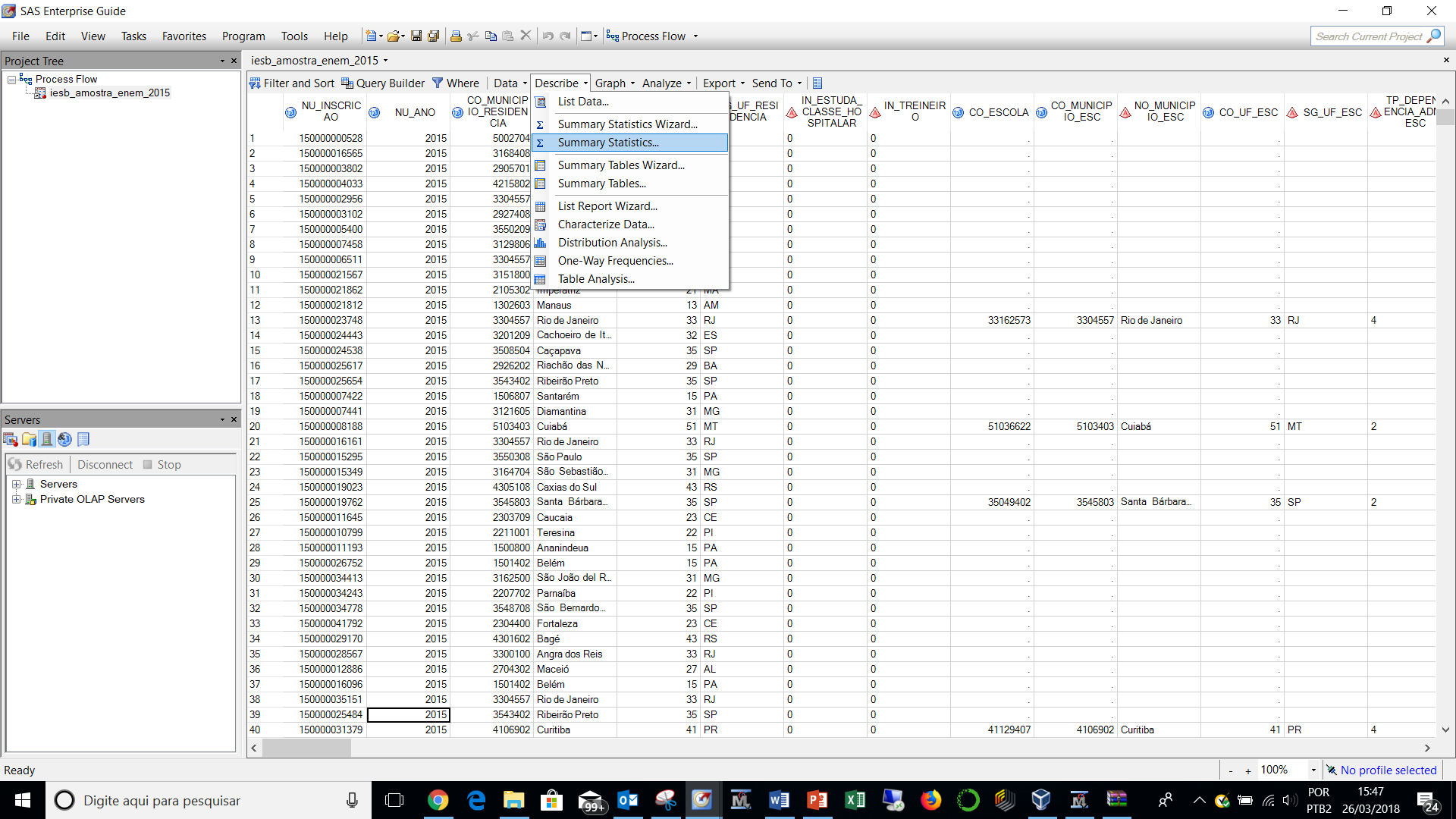


Explore as demais opções e coloque títulos, linhas verticas, nas etapas posteriores a definição do gráfico.

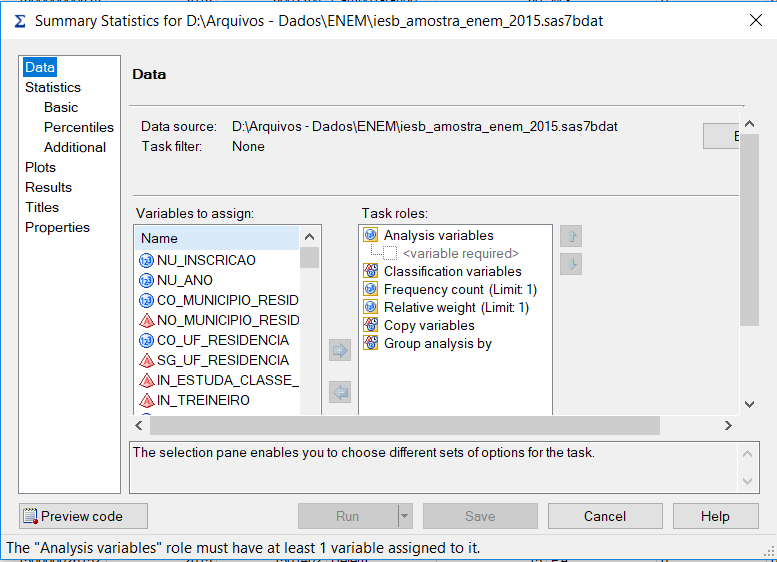
* 1. Explore a opção Graph e desenvolva novos gráficos para variáveis categóricas e numéricas;

Resultados:

1. Utilizando o Dataset SAS “dados\_enem\_amostra\_aula\_selec” vamos calcular algumas estatísticas.
   1. Selecione a opção “Summary Statistics”



* 1. Esta opção apresentará as seguintes opções:



Explore esta *procedure*, conforme orientação do professor.

* 1. Selecione o componente Summary Statistics.
     1. Faça uma análise da variável NU\_Nota\_MT, em nível Brasil. Calcule estatísticas de resumo, o histograma e coloque títulos em suas análises. O que você pode afirmar dos candidatos em relação a nota de matemática?

Resultados: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Faça uma análise da variável NU\_Nota\_MT, agora por Unidade da Federação da escola (SG\_UF\_RESIDENCIA). Calcule estatísticas de resumo, o histograma e coloque títulos em suas análises. O que você pode afirmar dos candidatos em relação a nota de matemática nos estados? Qual o estado com melhor nota? Qual o estado com pior nota?

Resultados: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Repita a sua análise para as variáveis NU\_Nota\_CH, NU\_Nota\_CN, NU\_Nota\_LC, NU\_Nota\_Redacao e Nota\_Media, Brasil e por Unidade da Federação da escola (SG\_UF\_RESIDENCIA). Calcule estatísticas de resumo, o histograma e coloque títulos em suas análises. O que você pode afirmar dos candidatos em relação as notas no Brasil e nos estados? Qual o estado com melhores notas no geral? Qual o estado com as piores notas no geral?

Resultados: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Substitua o arquivo dados\_enem\_amostra\_aula\_selec por um dos arquivos dados\_enem\_2017, dados\_enem\_2016 e/ou dados\_enem\_2015 e reexecute todas as suas análises.