

Introdução (RÁPIDA) ao SAS

O que é o SAS?

- Desenvolvido no início da década de 70 na North Carolina State University, Raleigh, NC
- No início, o objetivo era analisar estatisticamente experimentos agrícolas e para uso em investigação científica.
- SAS – Statistical Analysis System.
- Faturamento em 2020 de 3 bilhões de dólares.
- **James Goodnight** (404º mais rico do mundo: patrimônio de 6,5 bilhões de dólares, em março de 2022, segundo a Revista Forbes).
- **John Sall** (co-fundador e vice presidente, 956º mais rico do mundo: patrimônio de 3,2 bilhões de dólares em 2022 – Forbes).
- Mais de 12.554 empregados em 145 países.
- 91 das 100 maiores companhias no mundo usam SAS.
- Todas as 100 maiores universidades no mundo usam SAS (QS Top Universities).

SAS



James Goodnight
Fundador e CEO

SOFTWARE REVENUE BY INDUSTRY



banking
30%



capital markets
2%



communications
5%



education
2%



energy & utilities
2%



government
17%



health care
5%



insurance
11%



life sciences
8%



manufacturing
4%



retail
4%



services
10%

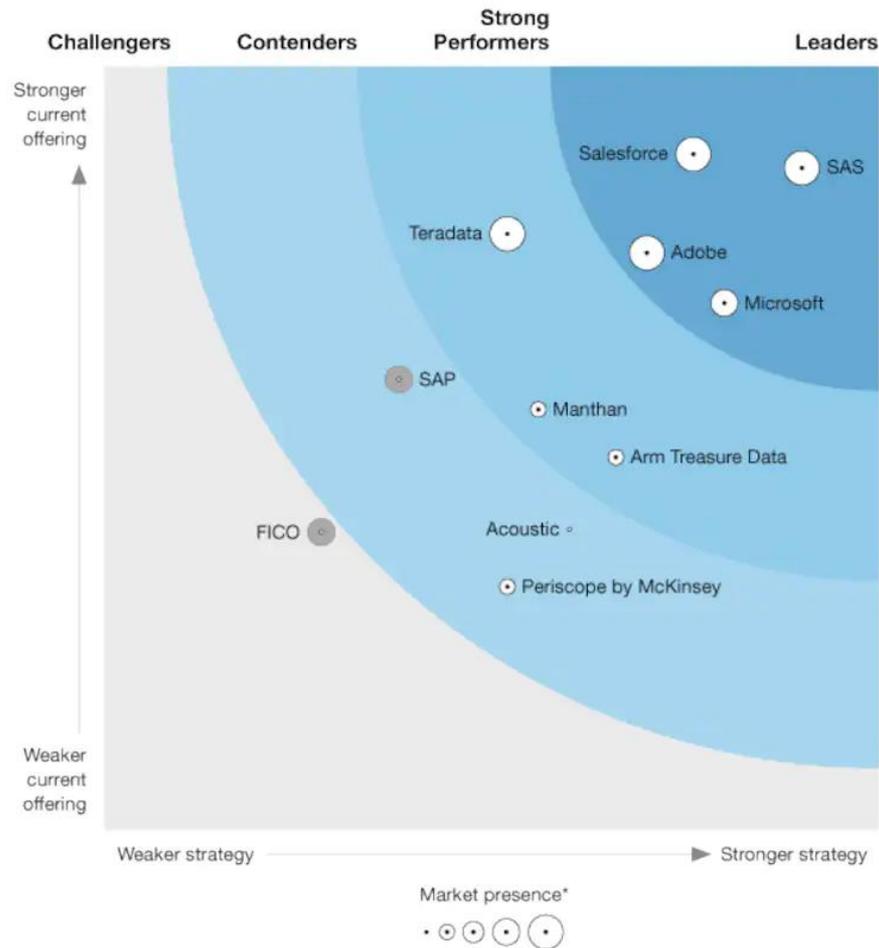
Liderança do SAS

- Especialista em análise de grande quantidade de dados (data mining big data).
- Dados meteorológicos: mais de 20.000 estações no mundo coletando dados a cada 15 minutos.
- Analítica: ciência da análise de dados.
- Primeira empresa em analítica preditiva e mineração de dados, entre as 9 maiores empresas mundiais (Forrester, 2013).
- Em 2008 adquiriu a empresa TERAGRAM, especializada em **mineração de texto** (clientes: Yahoo!, Banco Mundial, CNN, Forbes, NY Times, Washington Post, etc.)
- Em 2010 adquiriu a empresa Assetlink, líder no gerenciamento integrado de mercado (marketing).

THE FORRESTER WAVE™

Customer Analytics Technologies

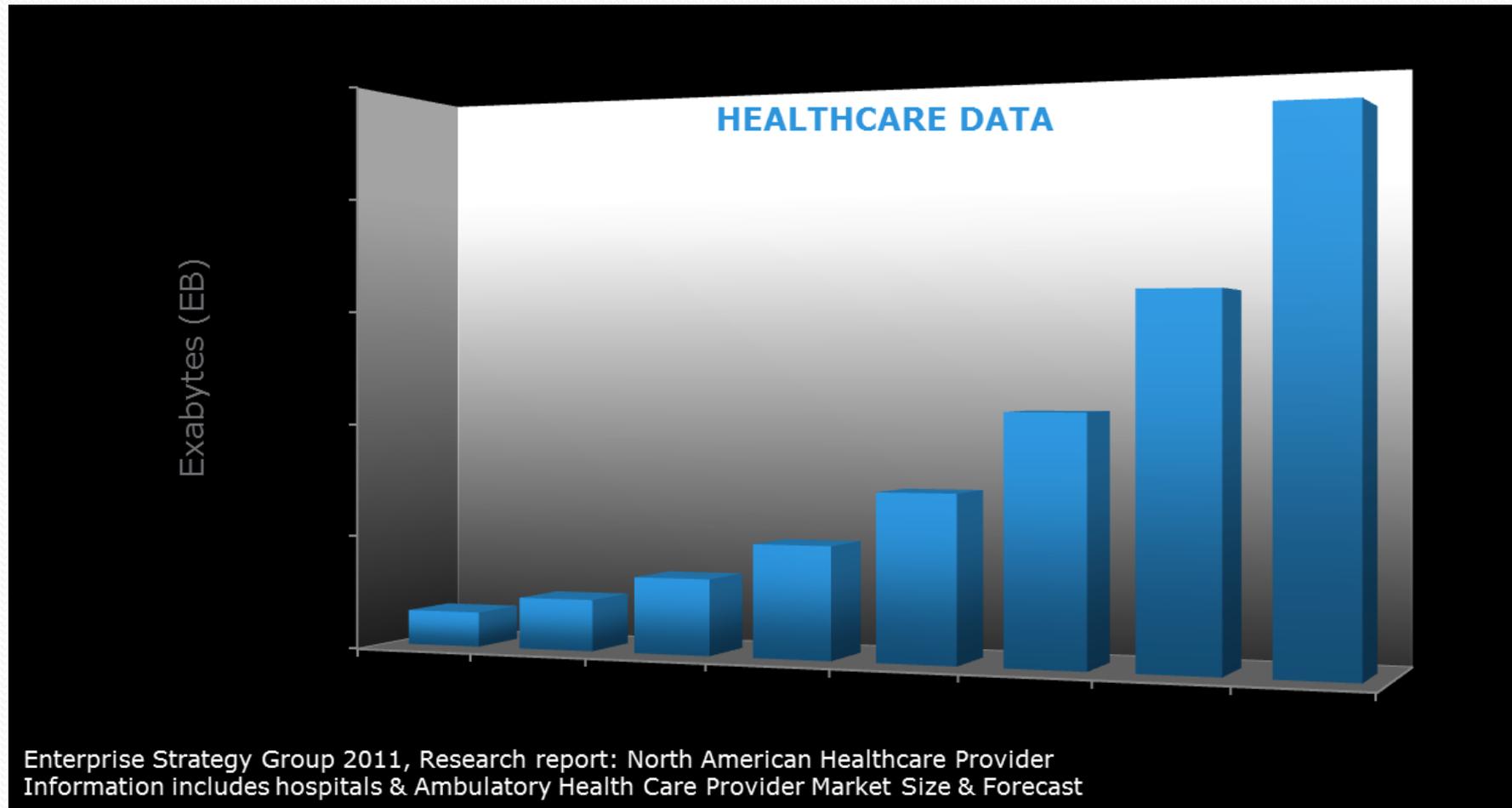
Q3 2020



*A gray bubble indicates a nonparticipating vendor.

Liderança em Analítica. 2021

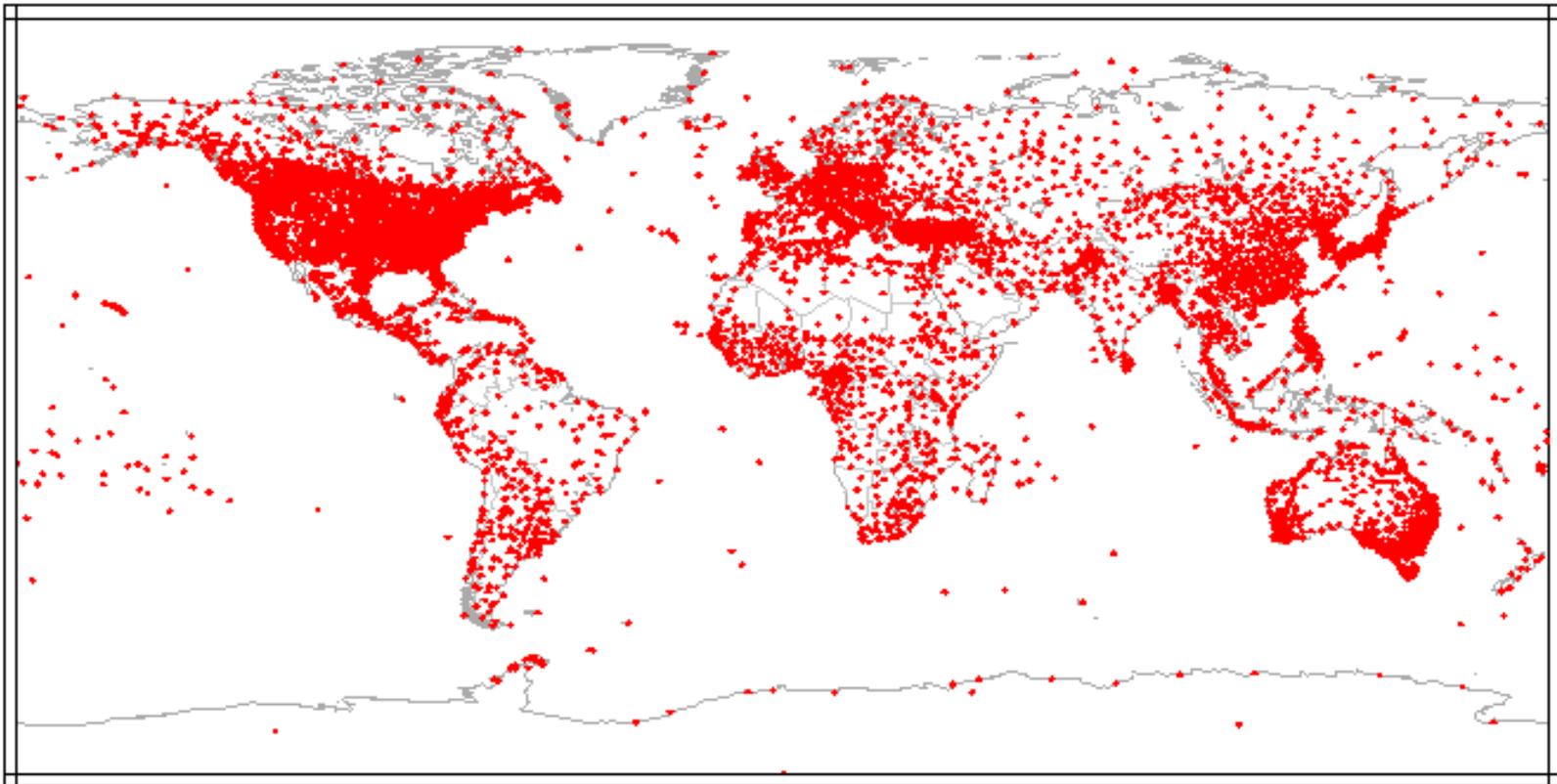
Crescimento do estoque de dados na área da saúde.



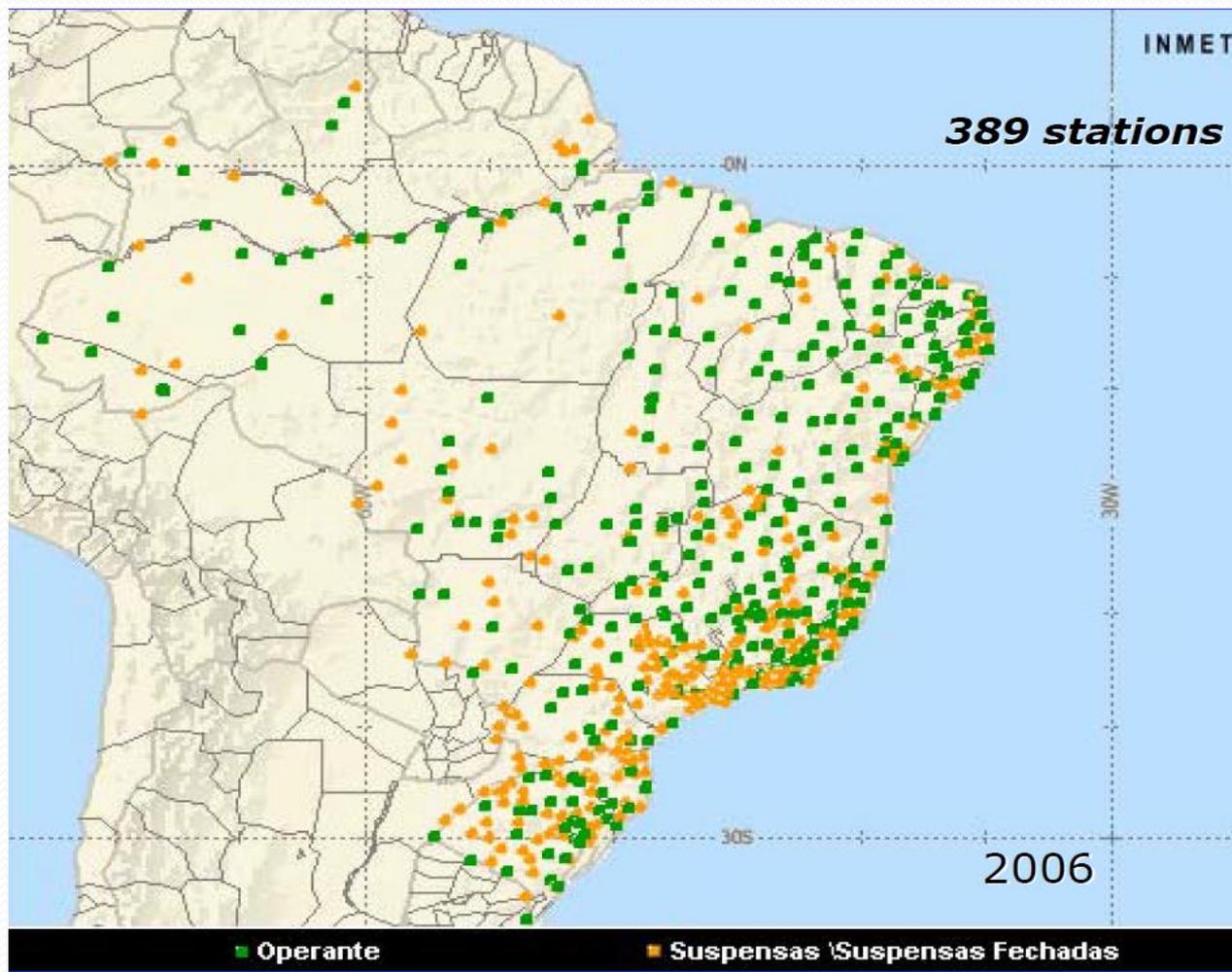
Rede de estações meteorológicas sob a supervisão da WMO.

- Mais de 10.000 na superfície terrestre.
- Mais de 1.000 estações aéreas fixas.
- Mais de 7.000 navios com estações meteorológicas.
- Mais de 1.000 bóias marítimas.
- Centenas de radares meteorológicos.
- Mais de 3.000 aeronaves comerciais equipadas com sistemas de monitoramento climático.

Estações meteorológicas fixas (WMO)



Estações meteorológicas fixas do INMET



As estações automáticas coletam dados a cada 15 minutos:
temperatura do ar,
precipitação, vento,
umidade relativa do ar,
radiação solar.

Produtos SAS

- Base – SAS – procedimentos básicos para o manejo de dados.
- SAS/Stat: análise estatística
- SAS/Graph: apresentação gráfica de qualidade.
- SAS/OR : pesquisa operacional
- SAS/ETS: econometria e séries temporais.
- SAS/IML: linguagem matricial interativa
- SAS/AF: para desenvolvimento de menus e interfaces.
- SAS/QC: controle de qualidade

Documentação

- <http://www.sas.com>
- **Versão que estamos usando: 9.4**
- <http://support.sas.com/documentation/>:
documentação geral do SAS para todas as versões.

Estrutura básica do SAS

- 3 componentes principais na maioria dos programas:
 - Comandos DATA
 - DADOS
 - Comandos PROC (procedure).

Comandos DATA

- Lê os dados de fontes internas ou externas, manipulam os dados e combinam com outros dados.
- Se presta para preparar os dados para serem usados por um ou mais procedimentos.

- Exemplo:

- ```
DATA SOLO;
INPUT LOCAL $ PONTO PROF MO;
C=MO*0.58; /* Transforma matéria orgânica em
carbono*/
CO2 = C*44/12; /* Transforma carbono em CO2, gás do
efeito estufa */
DATALINES;
```

.....

# Comandos PROC

- Realiza as análises dos dados e produz a listagem na janela OUTPUT.
- Exemplos de Proc:
  - PROC PRINT
  - PROC SORT
  - PROC MEANS
  - PROC FREQ
  - PROC UNIVARIATE
  - PROC GLM

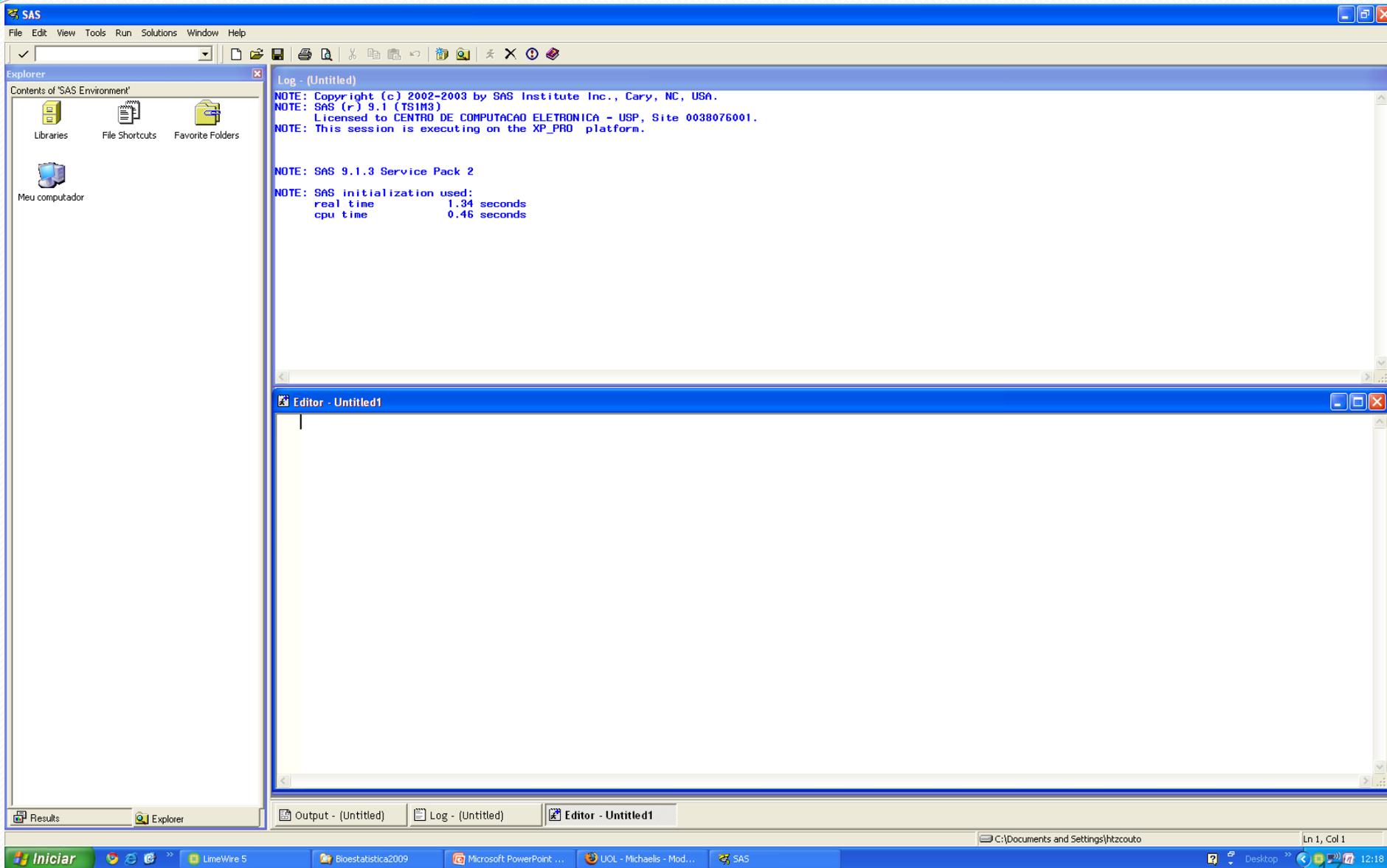
# Conceitos e Regras do SAS

- O nome SAS deve ter no máximo 32 caracteres, composto de letras, números e do caractere sublinhado (  )
- Não deve começar por número.
- Não deve ter espaços em branco.
- Cada sentença SAS deve terminar com o ponto e vírgula (;)
- Linhas que começam com \* e terminam com o ; são tratadas como comentário.
- Os comentários também podem ser colocados entre /\* e \*/.
- RUN; é o comando usado para executar o programa.

# Janelas

- 3 janelas principais
- EDITOR usada para criar, editar e executar um programa SAS.
- LOG relata o progresso do programa SAS, mostra os erros (vermelho) e os avisos (verde).
- OUTPUT mostra a saída de um programa SAS. Pode-se copiar, salvar ou imprimir a listagem do programa SAS.

# Principais janelas do SAS



# Entrada de dados para análise

- 3 modos de entrada de dados no SAS:
  - Ler dados digitados ou colados na janela EDITOR
  - Importar dados do Excel.
  - Outros procedimentos, como importar arquivos .txt, Access, DBF, etc.

# Ler dados digitados ou colados no EDITOR

**DATA ALUNOS;**

**INPUT NOME \$ ALTURA DT\_NAS DDMMYY10.;**

**HOJE = DATE();**

**DIAS = HOJE – DT\_NAS;**

**IDADE\_ANOS=ROUND ((DIAS/364.5),0.1);**

**DATALINES;**

JOANA 1.87 12/03/2010

PEDRO 1.67 01/05/2009

MARIA 1.68 03/07/1999

;

**PROC PRINT DATA=ALUNOS;**

**FORMAT DT\_NAS DDMMYY10.;**

**RUN;**

# Importar dados do Excel

- PROC IMPORT OUT=WORK.Ao
- DATAFILE='F:DADOS1.XLSX'
- DBMS=XLSX REPLACE;
- RUN;
- Se os dados não estiverem na Planilha usar outro comando: SHEET='Planilha3'; por exemplo

**Os dados estão na primeira planilha, qualquer que seja o nome dela.**

# Saída SAS

- A listagem (resultados das análises) pode ser gravada em HTML (**H**yper**T**ext**M**arkup**L**anguage); RTF (**R**ich**T**ext**F**ormat – pode salvar e editar no Microsoft WORD) e PDF (**P**ortable**D**ocument**F**ormat), entre outros formatos.
- Na janela OUTPUT do SAS: o conteúdo pode ser copiado e colado, por exemplo no Microsoft WORD, ou impresso diretamente.

# Programa SAS para gravar arquivo em PDF (Adobe Acrobat)

```
PROC IMPORT OUT= WORK.A1
 DATAFILE= "C:\INVENTÁRIO2016\DAEXC.xls"
 DBMS=EXCEL REPLACE;
 SHEET="Plan1$";
 GETNAMES=YES;
 MIXED=NO;
 SCANTEXT=YES;
 USEDATE=YES;
 SCANTIME=YES;
RUN;
ODS PDF FILE='C:\RESULTADO.PDF';
PROC PRINT DATA=A1;
RUN;
ODS PDF CLOSE;
```

# Alguns comandos SAS

- LENGTH
- IF... THEN...;
- DELETE
- AND OR
- EQ, NE, GE, LE, GT, LT

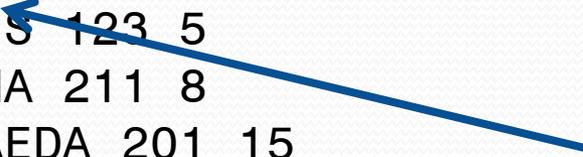
# Programa SAS:

## Uso do comando LENGTH

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Uso de comando condicional: IF ... THEN ...;

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF LOCAL ='RCLARO' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇOIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```



Caractere  
entre  
aspas

# Uso de comando condicional:

## IF ... THEN ...:

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF IDADE=7 THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇÓIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

Número sem aspas

# Uso de comando condicional: AND (E)

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF IDADE=7 AND LOCAL='BAURÚ' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇÓIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Uso de comando condicional: OR (OU)

```
DATA A;
LENGTH LOCAL ESPECIE $12.;
INPUT LOCAL $ ESPECIE $ VOLUME IDADE;
IF LOCAL='LENÇÓIS' OR LOCAL='BAURÚ' THEN DELETE;
DATALINES;
RCLARO E.GRANDIS 123 5
RCLARO E.SALIGNA 211 8
PIRACICABA P.TAEDA 201 15
PIRACICABA P.ELLIOTTII 98 8
BAURÚ E.GRANDIS 321 7
LENÇÓIS E.UROPHYLLA 303 7
;;;
PROC PRINT DATA=A;
RUN;
```

# Comandos de comparação

- EQ = IGUAL A
- NE = DIFERENTE DE
- GE = MAIOR OU IGUAL A
- LE = MENOR OU IGUAL A
- GT = MAIOR QUE
- LT = MENOR QUE

# Programa SAS: modificação dos dados.

```
DATA SOBREV;
 INPUT CLONE $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: EQ (Igual a )

```
DATA SOBREV;
 INPUT CLONE $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF TAXAMORT EQ 0 THEN DELETE;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: NE (Diferente de)

```
DATA SOBREV;
 INPUT CLONE $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF ESP NE 'A02' THEN DELETE;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: GE (Maior ou igual a)

```
DATA SOBREV;
 INPUT CLONE $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF IDADE GE 5 ;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Programa SAS: LE (Menor ou igual a)

```
DATA SOBREV;
 INPUT CLONE $ IDADE VIVOS ;
 TAXAMORT= ((25-VIVOS)/25)*100;
 IF IDADE LE 2 THEN VIVOS=25 ;
DATALINES;
A01 0 25
A01 1 25
A01 2 21
A01 3 18
A01 4 11
A01 5 7
A01 6 3
A02 0 25
A02 1 25
A02 2 25
A02 3 14
A02 4 6
A02 5 1
A02 6 0
;;;
PROC PRINT DATA=SOBREV;
RUN;
```

# Exercício

- A) Fazer um programa SAS para selecionar apenas as idades maiores que 4 (Use o comando GT).
- B) Fazer um programa SAS para selecionar apenas as idades cujos números de sobreviventes sejam menores que 15 (Use o comando LT).

Obrigado e até a  
próxima semana!!!