SAS Enterprise Guide
8th Annual SAS® Summer Institute

by the University of Iowa SAS® User Group
August 16-17, 2015
Yelena Perkhounkova
elena-perkhounkova@uiowa.edu

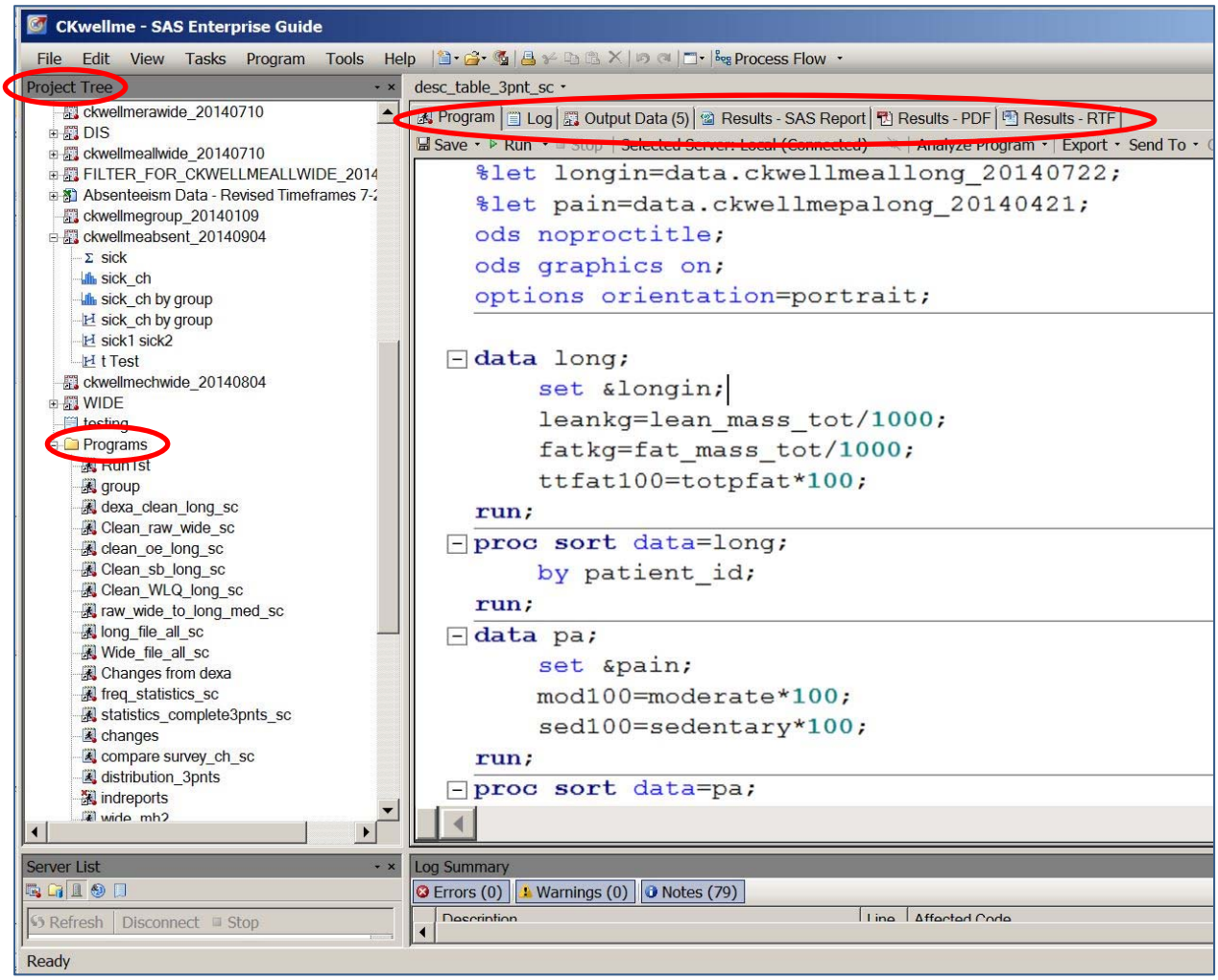
What is SAS Enterprise Guide?

A point-and-click graphical interface to SAS that provides

- easy access to data sources/export capabilities
- ready-to-use ***tasks*** to make reporting & analytics more available
- programming interface
- organizational structure to **projects** that include
  - SAS programs
  - References to data and associated tasks
  - Results in various formats
  - Logs
  - Relationships among the items above

SAS EG project shows references to data, tasks, and programs in the project tree.

The program, log, output data, and results are available via tabs in the main pane.

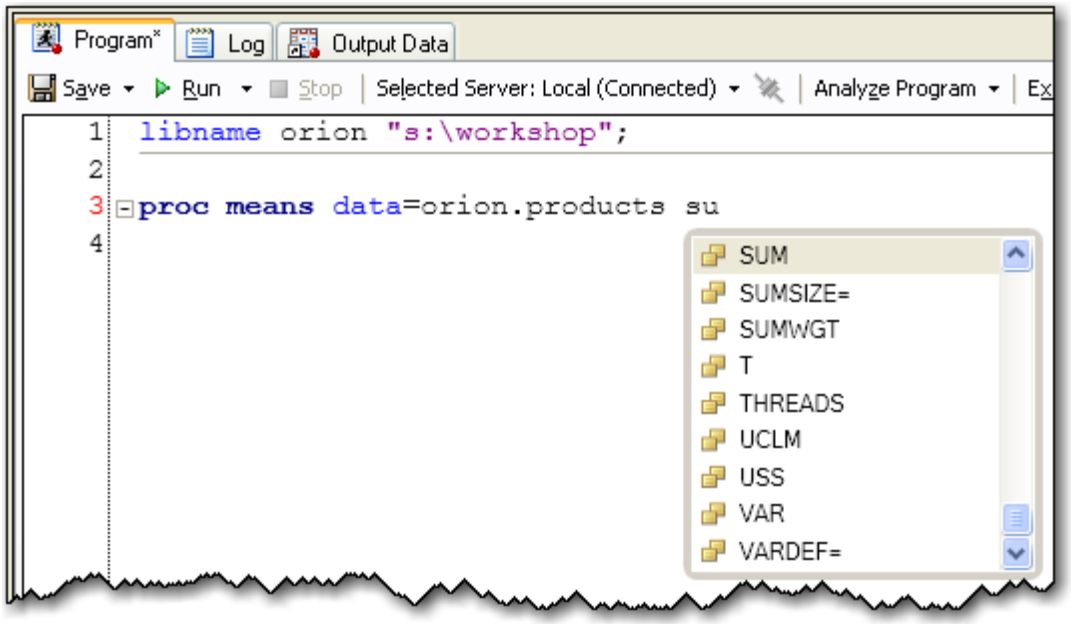


Why SAS Enterprise Guide?

- Important: you can write code in SAS EG just like in any SAS environment!
- However, if your data are already clean and properly set up, you don't need to write a single line of code in SAS EG to do a lot of different analyses.

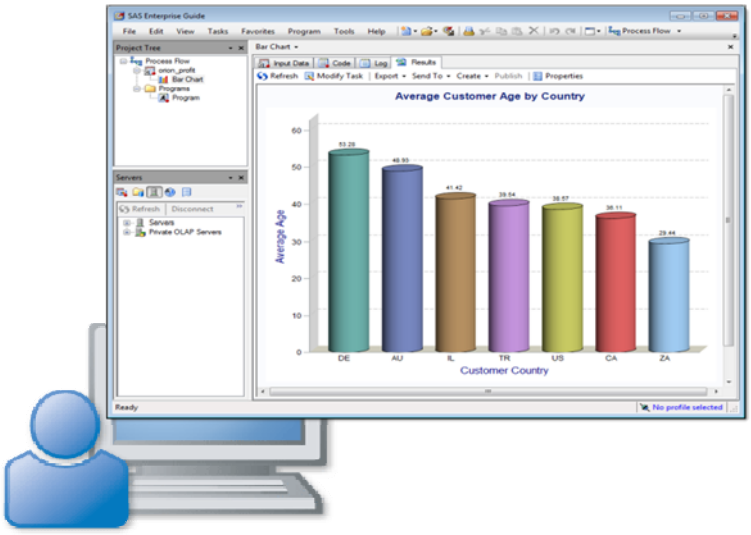
Enterprise Guide Program Editor provides convenient tools for writing code

- autocomplete
- dynamic syntax tooltips
- formatting programs to provide consistent spacing
- analyzing program flow



How you can access SAS EG

- Virtual Desktop (7.1)
- SAS installed on your PC (7.1)



Additional information

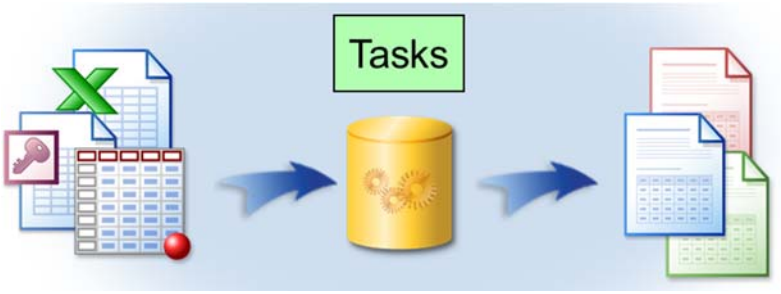
- Getting Started with SAS Enterprise Guide Tutorial

  http://support.sas.com/documentation/onlinedoc/guide/tut71/en/

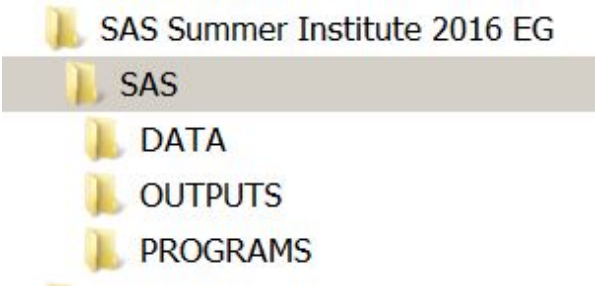
  http://support.sas.com/eguide

I will show today how to use SAS EG to

- Create and save projects
- Use SAS EG tasks to
  - Add SAS data to the project/export data outside SAS
  - Create subsets of data
  - Summarize data
  - Plot data
  - Run simple statistical analysis
- Add programs to the projects



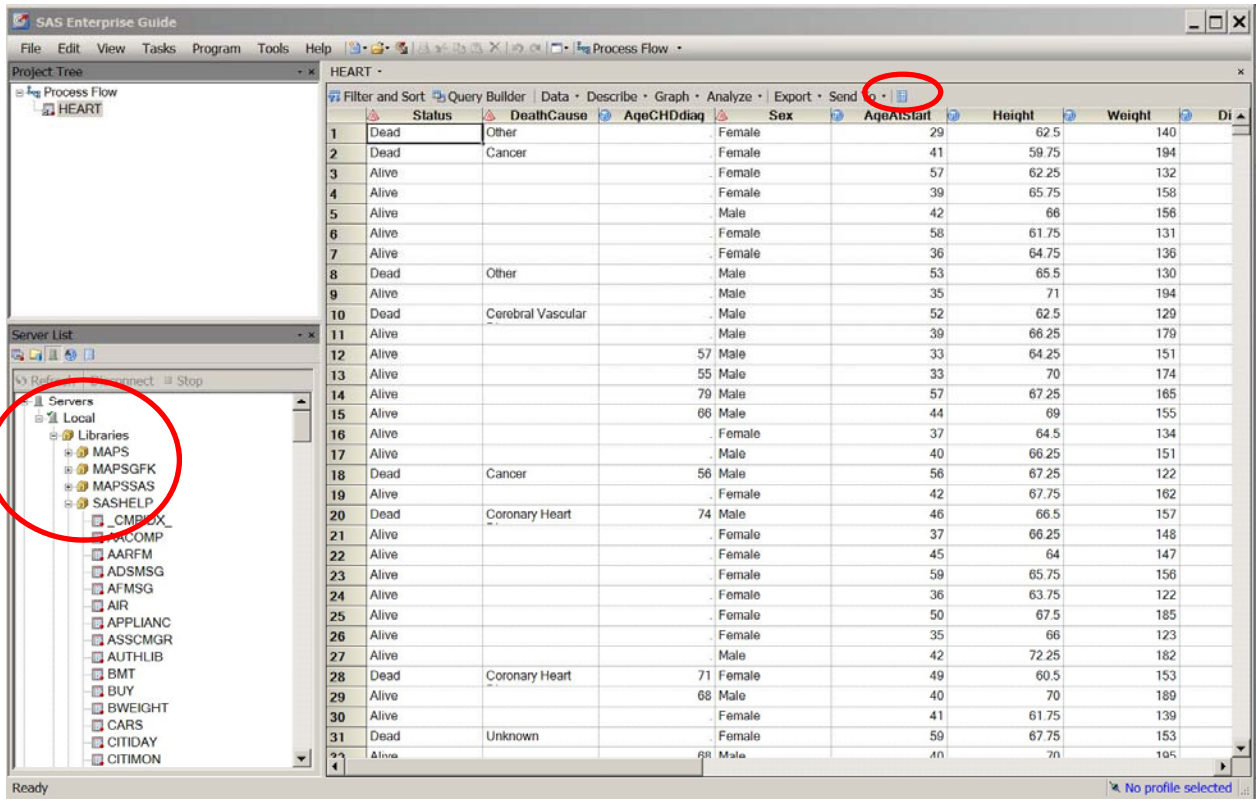
Start by creating convenient folder structure on H-Drive



Open a new project in EG.

Under Server List, go **to Servers→Local→Libraries→SASHELP**

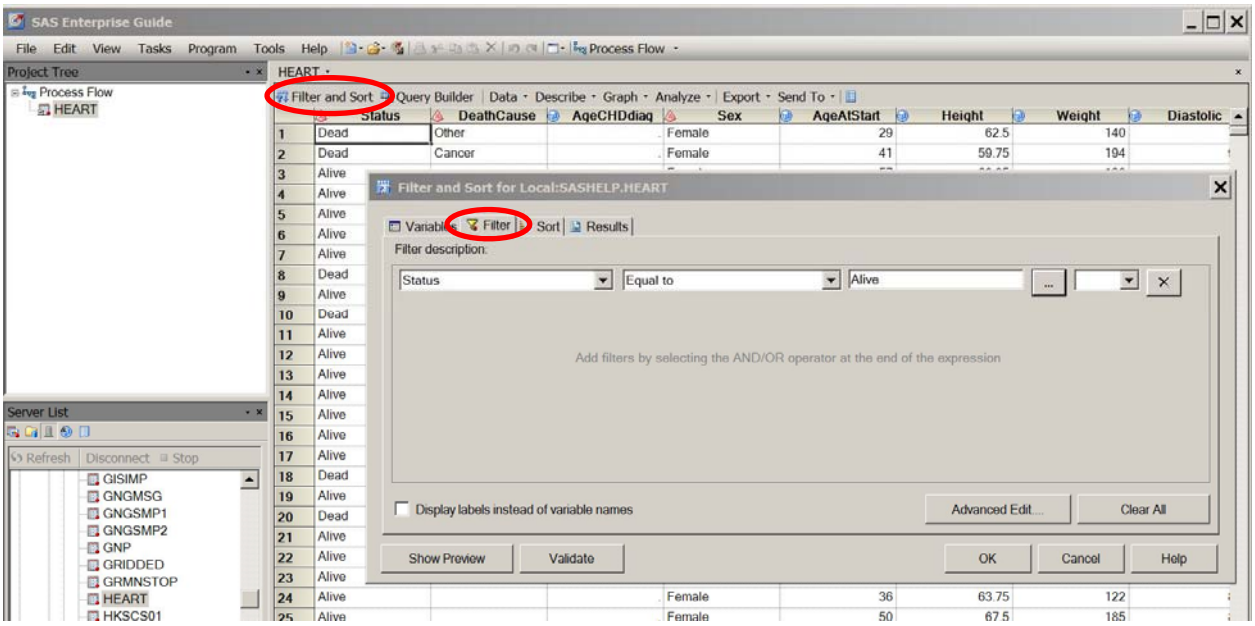
Open HEART dataset



To explore file properties right-click on file in project tree (or click on “properties” tab)

To select only living people use “Filter and Sort” task

- Select “Filter and Sort” tab
- Select all variables
- Choose filter (**Status** equal to “Alive”)
- Click “OK”



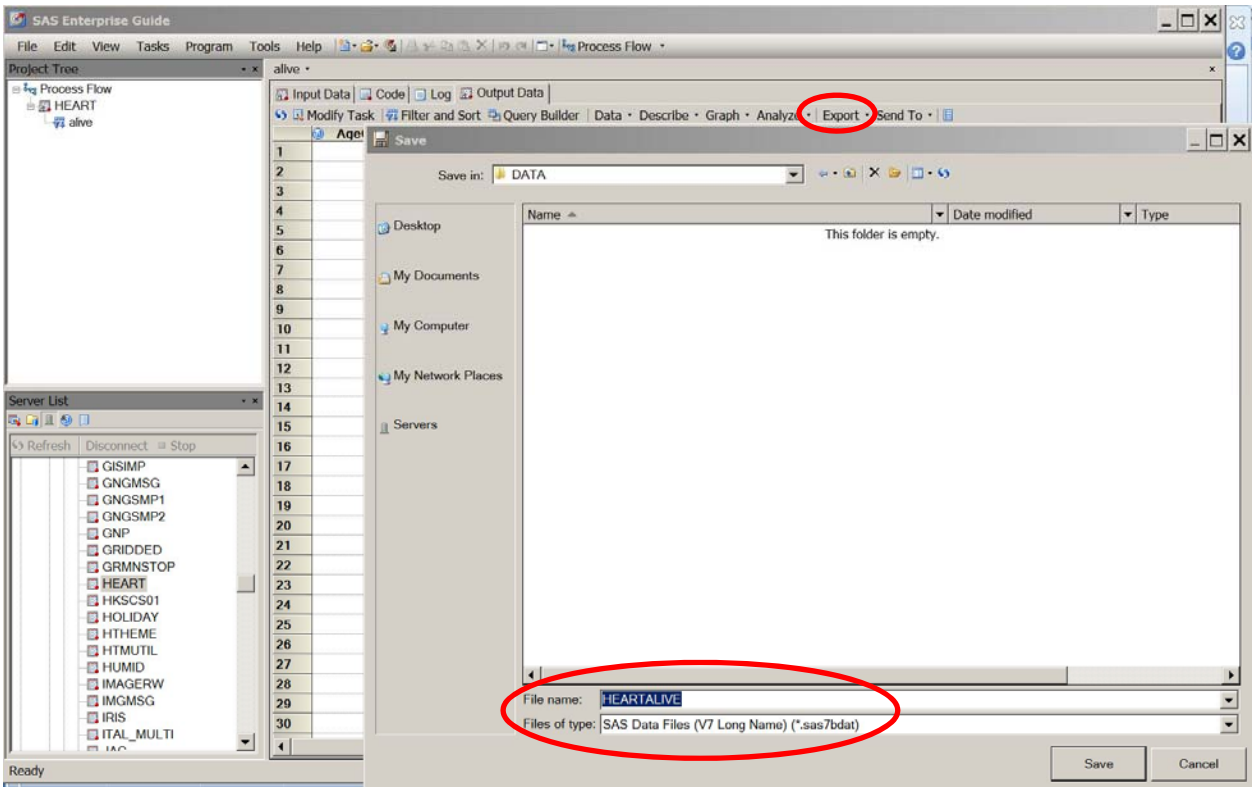
Name new “Filter and Sort” task “alive”

To edit a filter

- Select it
- Select “Modify Task”
- Remove **Status, Deathcause,** and **AgeAtDeath** from list of selected variables
- Click “OK”

To save new data to H:\your project\SAS\DATA

- Select the “export” tab
- Choose “Export ‘filename’”
- Save as HEARTALIVE



Save your project to H:\your project\SAS\PROGRAMS

Name the project “SASEG16”

Close the project

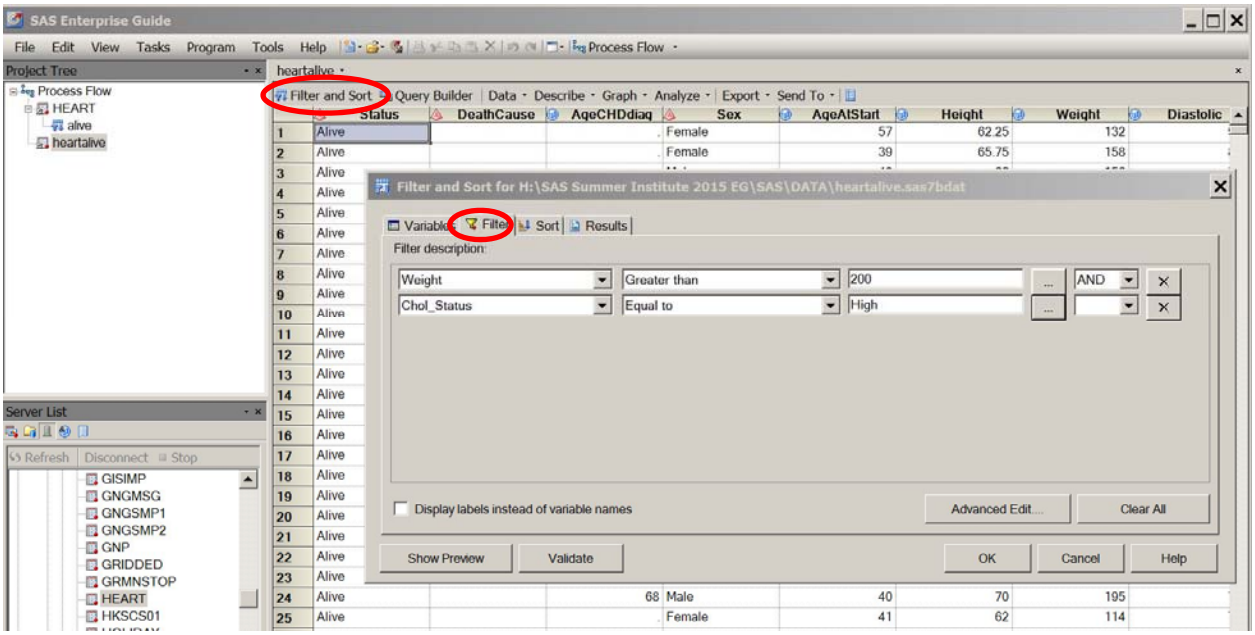
Reopen your project

To open existing SAS file into project

- Select **File → Open → Data**
- Browse, select & open file HEARTALIVE

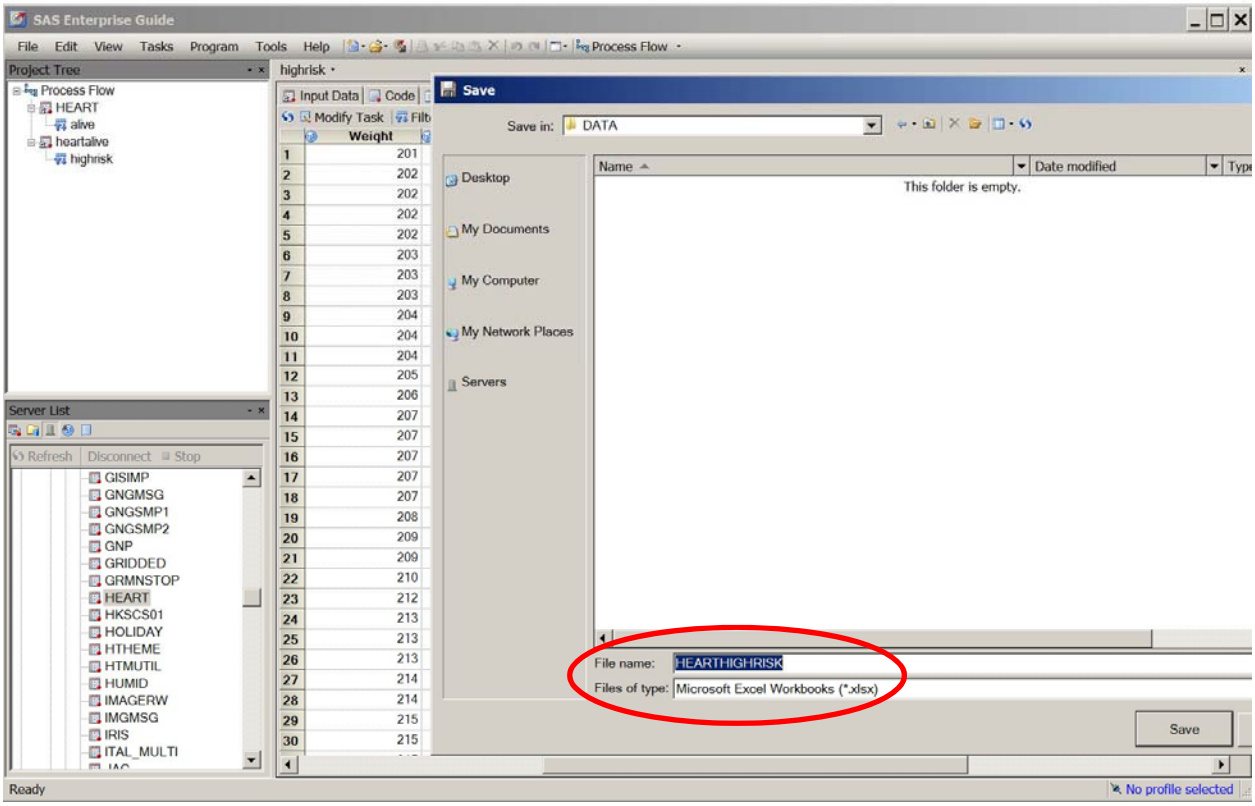
To select people > 200 lbs with high cholesterol status create a new “Filter and Sort” task

- Select “Filter and Sort” tab
- Select all variables
- Choose filters (**Weight** > 200 AND **Chol_status** equal to “High”)
- Sort by **Weight**
- Name task “highrisk” in “Results” tab
- Click “OK”



To save data in Excel format use “Export” task

- Select “Export” tab
- Choose “Export ‘filename’”
- Choose Excel file type
- Name file as HEARTHIGHRISK
- Save



To open non-SAS data file (e.g. Excel) into the project

- Select **File→Import data**

  Browse, select HEARTHIGHRISK.xlsx file, open, follow instructions:

  1. "Specify the data": click next
  2. "Select Data Source": choose worksheet, select "rename columns to comply with SAS naming", click next
  3. "Define field attributes": click next
  4. "Advanced options": select "remove characters that can cause transmission errors from text-based data files", click finish

- Export to your DATA folder as SAS data file with the name "HEARTHIGHRISKfromExcel"
- Open HEARTHIGHRISKfromExcel into your project
- Compare properties of HEARTALIVE and "HEARTHIGHRISKfromExcel files
- Double-click on HEARTALIVE to make it active (we will use it from now on)

To calculate summary statistics for continuous variables use “Summary statistics” task (PROC MEANS)

- Select **“Describe” tab→Summary Statistics**
  - Data

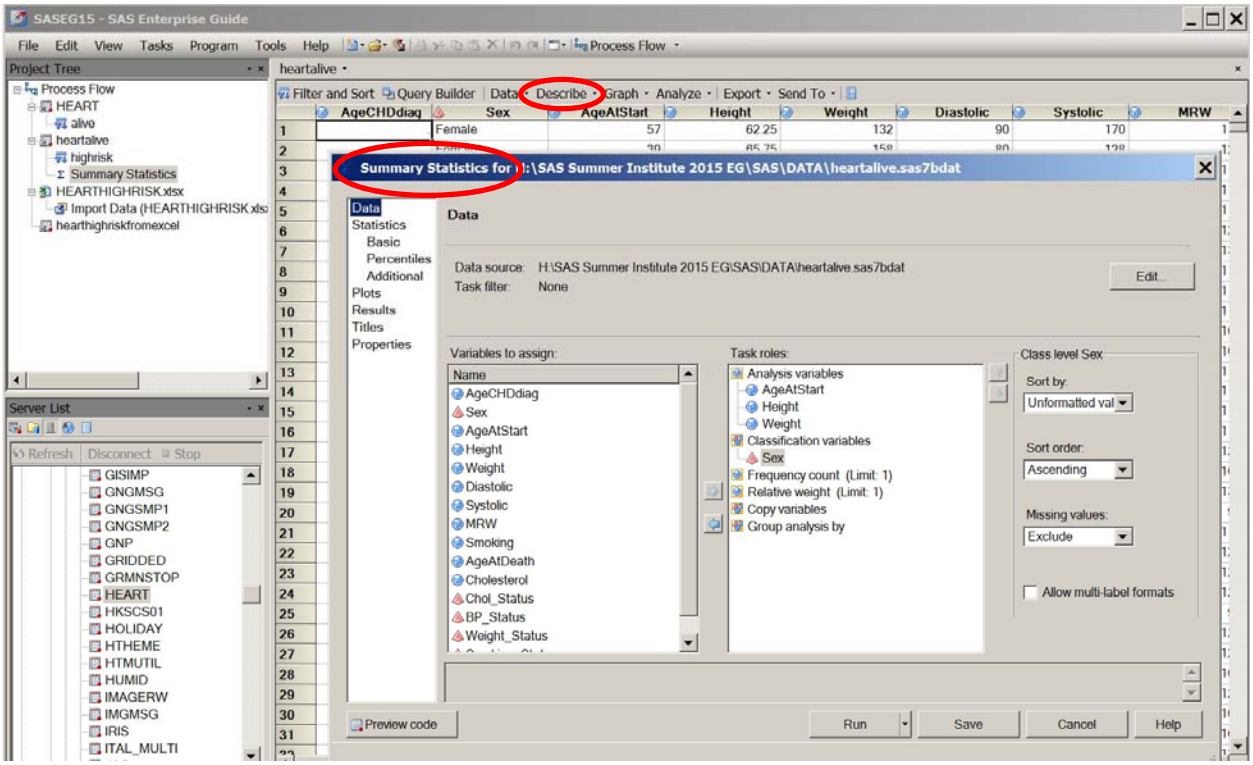
    Analysis variables: **AgeatStart**, **Height**, **Weight**

    Classification variable: **Sex**
  - Statistics

    Basic: select mean, standard deviation, min, max, number of observations

    Percentiles: select median

    Additional: select confidence limits of the mean
  - Plots: Select histogram and box-and-whisker plot
  - Titles: Change title to “Summary Statistics for Age, Height, and Weight”



- Click on “Run”
- Examine output
- Examine code
- Name task “age height weight”

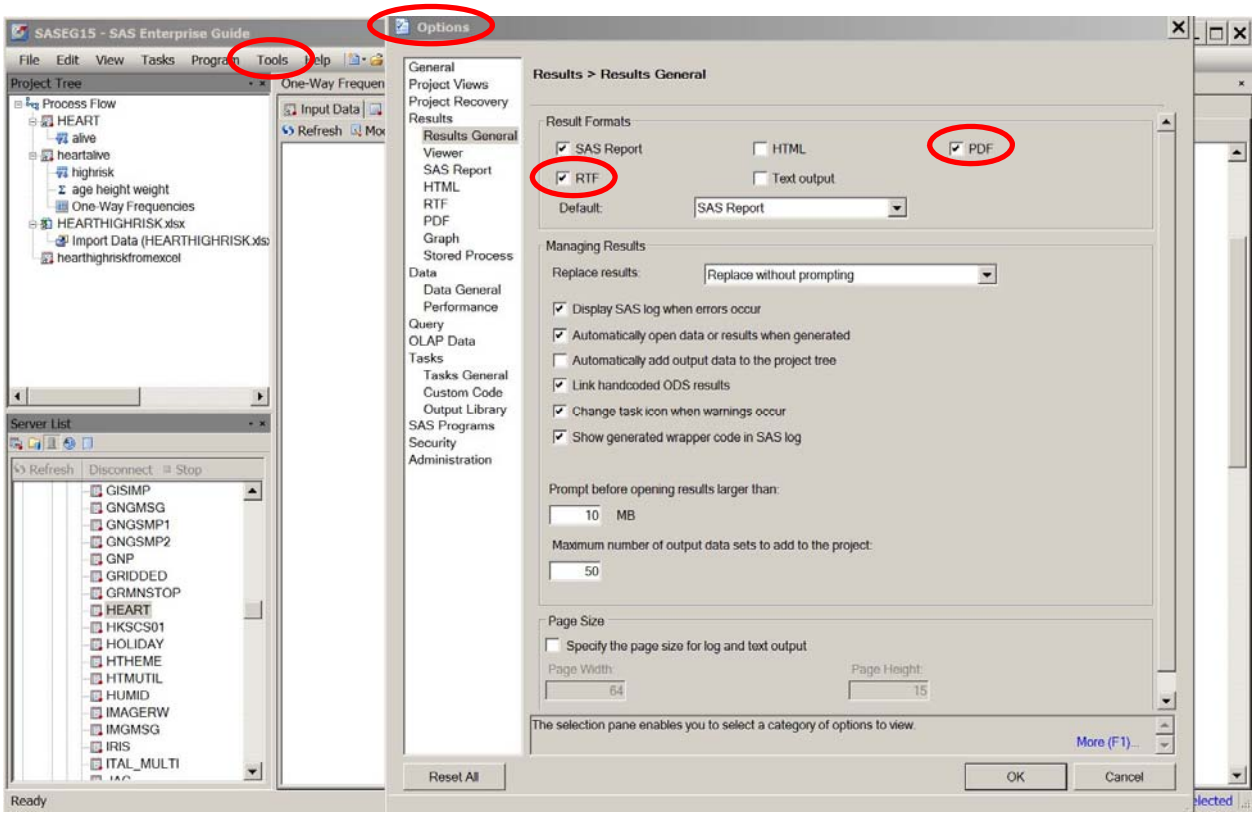
To summarize categorical variables "One-Way Frequencies" task (PROC FREQ)

- Double-click on HEARTALIVE data
- Select **"Describe" tab→One-Way Frequencies**
  - Data

    Analysis variables: **Chol_status, BP_Status, Weight_Status, Smoking_Status**
  - Plots: Select vertical bar chart
- Click on "Run"
- Examine output
- Check code
- Name task "statuses"

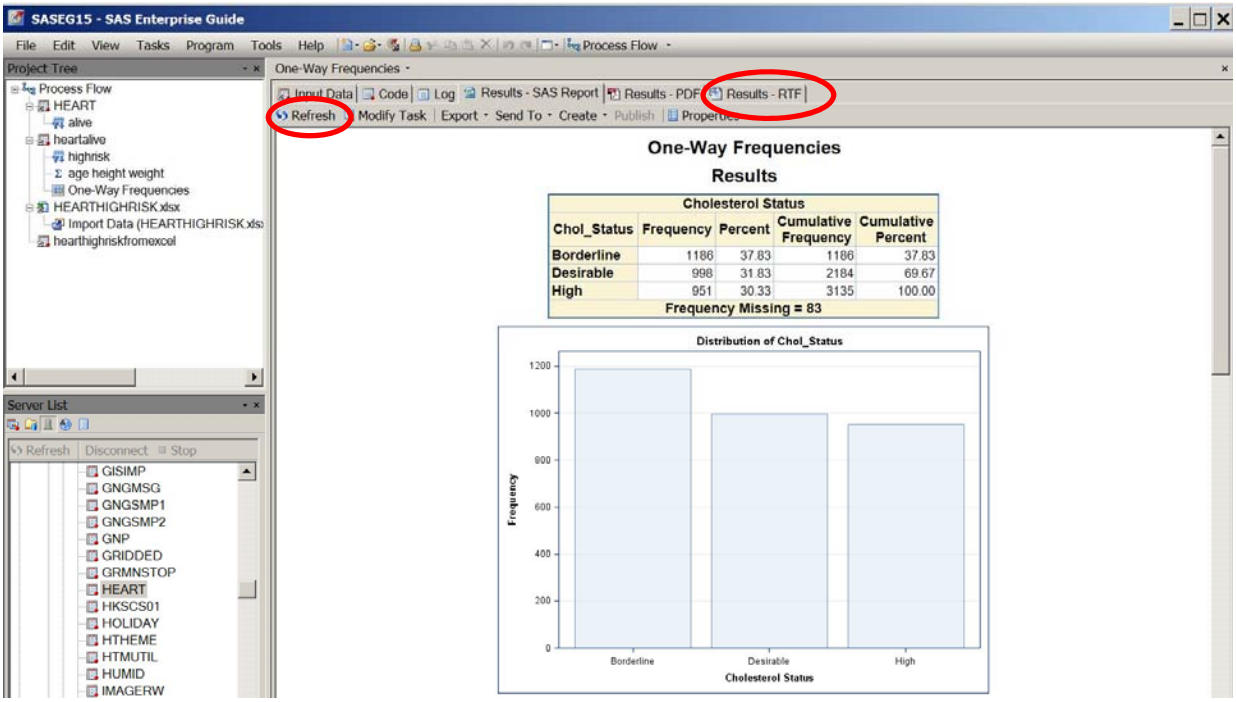
To create output in other formats modify SAS EG options

- Select **Tools→Options**
  - Results: Select PDF and RTF
- Click "OK"



To add PDF and RTF outputs to results

- Select “Refresh” tab
- Select “Results-RTF” tab
- Export the output to H:\your project \SAS\OUTPUTS



To examine distributions of continuous variables use “Distribution Analysis” task (PROC UNIVARIATE)

- Double-click on HEARTALIVE data
- Select **“Describe” tab→Distribution Analysis**
  - Data

    Analysis variables**: Diastolic**, **Systolic**
  - Distributions

    Normal: select “Normal”, “Suppress distribution tables”
  - Plots

    Appearance: select histogram plot and probability plots
  - Tables: select basic measures, extreme rows, moments, tests for normality
- Click on “Run”

To examine bivariate relationships between continuous variables use "Correlations" task (PROC CORR)

- Select **"Analyze" tab → Multivariate → Correlations**
  - Data

    Analysis variables: **Weight**

    Correlate with: **Diastolic**, **Systolic**
  - Results: Select "Create a scatter plot for each correlation pair"
- Click on "Run"
- Use "Modify Task" to run analysis by **Sex**

To compare two independent samples on a continuous normal variable use "t test" task (PROC TTEST)

- Select **"Analyze" tab → ANOVA → t Test**
  - t Test type: two sample
  - Data

    Classification variable: **Sex**, **Weight**

    Analysis variables: **Cholesterol**
  - Plots: select summary plot
- Click on "Run"

To compare two independent samples on a continuous non-normal variable use "Nonparametric One-Way ANOVA" task (PROC NPAR1WAY)

- Select **"Analyze" tab → ANOVA → Nonparametric One-Way ANOVA**
  - Data

    Independent variable: Sex

    Dependent variables: Systolic
- Analysis: uncheck all, but Wilcoxon
- When more than 2 groups, use Kruskal-Wallis test results

To investigate a bivariate relationship between two categorical variables use “Table analysis” task (PROC FREQ)

- Select **“Describe” tab→ Table Analysis**
  - Data

    Table variables: **Sex**, **Chol_status, BP_Status, Weight_Status, Smoking_Status**
  - Tables: define 4 tables to be generated (sex in columns, statuses in rows)
  - Cell Statistics: Row percentages, column percentages, cell frequencies
  - Table Statistics

    Association: check Chi-square tests
- Click on “Run”
- Examine output
- Check code
- Name task “statuses by sex”

To create a line plot use “Line plot” task

- Select **“Graph” tab → Line Plot**
  - Data

    Horizontal: **BP_status**

    Vertical: **Weight**
  - Click on “Run”
  - Examine output: what’s wrong?
- Select “Modify Task” tab
  - Data: For **weight** select “Summarize for each distinct horizontal value”, select function “Average”
  - Appearance

    Axes

    Horizontal axis: select Reverse Axis

    Vertical axis: type label “Average weight”, rotate 90°
- To create plots for males and females, select “Modify Task” tab
  - Select “Multiple line plots by group column”
  - Data

    Group: **Sex**
  - Appearance

    Plots: add symbols
  - Titles: Type “Weight by Blood Pressure Status, for Males and Females”
- Click on “Run”

To create a new SAS program in the project

- Select **File → New → Program**
- To create new variables, type (see hints below)

```
Data tempheart;
  set 'H:\your project\SAS\DATA\heartalive';
  sqrtsystolic=sqrt(systolic);
  lnsystolic=log(systolic);
  lnsystolic70=log(systolic-70);
run;
```

- Click on tab "Run"

Hints:

- Go to the DATA folder and select and copy the path to the folder, then paste in place of "*your project*"
- Fix the quotes around the path

Examine distributions for new variables and compare to the original systolic variable. How do they differ?

To clean data in SAS EG directly (self-learning)

- Select **Tools → Options**
  - Data General: Select "Use data in unprotected mode"
- Create a copy of your original data (export with a new name into DATA folder)
- Open this copy into your project
- Make changes to data
- WARNING –unselect "Use data in unprotected mode" after you are done cleaning.