



**THE
POWER
TO KNOW®**

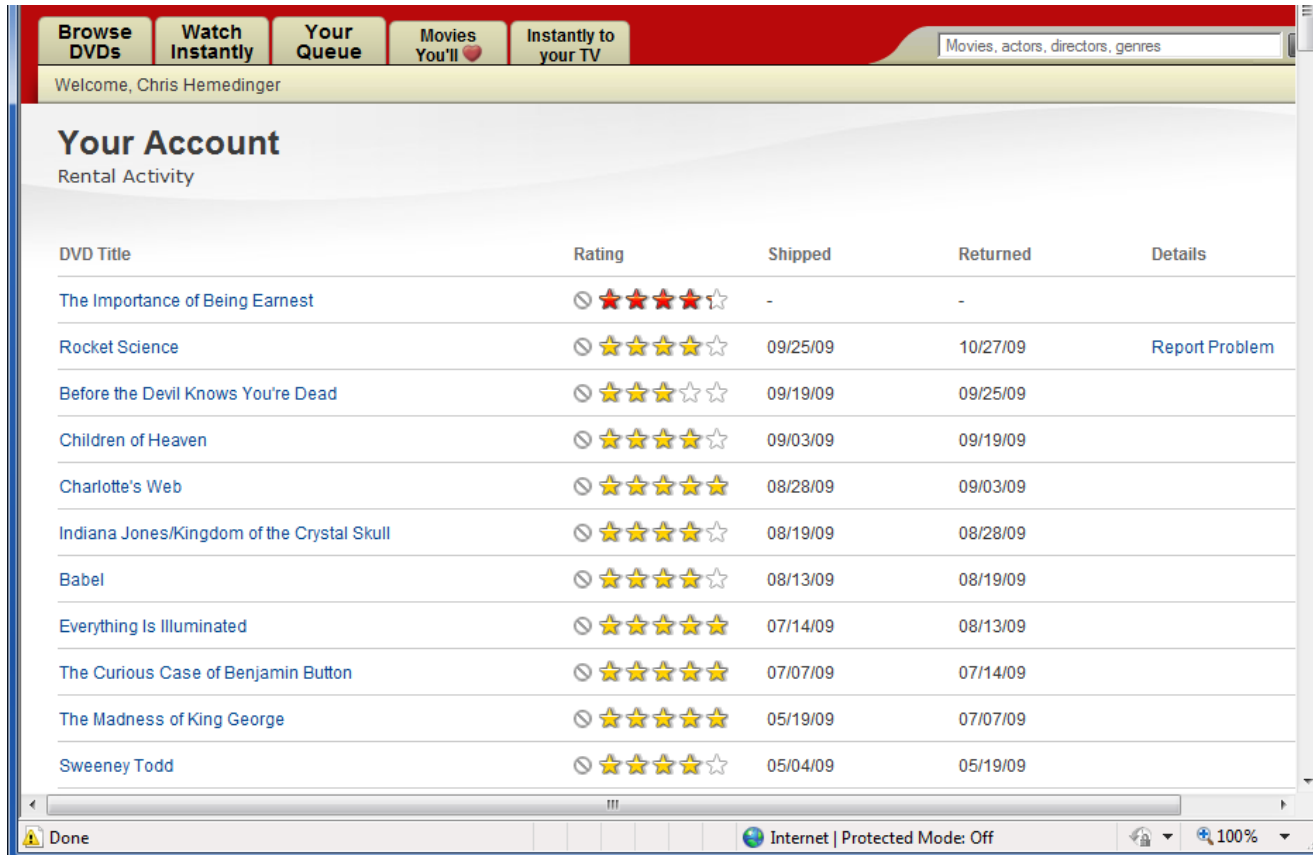
Upping Your SAS Game using SAS 9.2 and SAS Enterprise Guide

Chris Hemedinger
Senior Software Manager, SAS Institute

Agenda

- It's all about me
 - Using SAS to analyze my online movie rentals
 - Using SAS to analyze my Facebook friends
- Here's something for you!
 - Fast-forward: what are we doing for the SAS programmer?

Movie rentals: Plenty of raw data available



The screenshot shows a web browser window displaying a movie rental website. The top navigation bar includes links for 'Browse DVDs', 'Watch Instantly', 'Your Queue', 'Movies You'll ❤️', and 'Instantly to your TV'. A search bar contains the text 'Movies, actors, directors, genres'. Below the navigation bar, a welcome message reads 'Welcome, Chris Hemedinger'. The main section is titled 'Your Account' with a sub-header 'Rental Activity'. It contains a table with columns for 'DVD Title', 'Rating', 'Shipped', 'Returned', and 'Details'. The table lists 10 movies with their respective ratings (represented by stars) and shipping/return dates. The browser's status bar at the bottom shows 'Done', 'Internet | Protected Mode: Off', and a zoom level of '100%'.

DVD Title	Rating	Shipped	Returned	Details
The Importance of Being Earnest	4.5 stars	-	-	
Rocket Science	4.5 stars	09/25/09	10/27/09	Report Problem
Before the Devil Knows You're Dead	4.0 stars	09/19/09	09/25/09	
Children of Heaven	4.5 stars	09/03/09	09/19/09	
Charlotte's Web	5.0 stars	08/28/09	09/03/09	
Indiana Jones/Kingdom of the Crystal Skull	4.5 stars	08/19/09	08/28/09	
Babel	4.5 stars	08/13/09	08/19/09	
Everything Is Illuminated	5.0 stars	07/14/09	08/13/09	
The Curious Case of Benjamin Button	5.0 stars	07/07/09	07/14/09	
The Madness of King George	5.0 stars	05/19/09	07/07/09	
Sweeney Todd	4.5 stars	05/04/09	05/19/09	

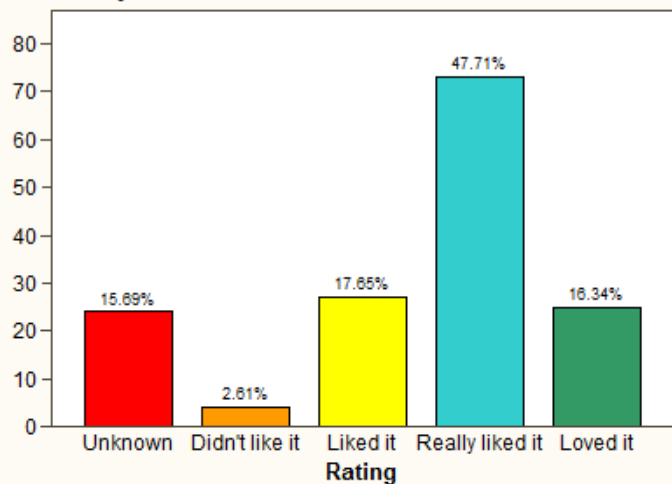
In Internet Explorer, right-click and select **Export to Microsoft Excel**

Netflix History Report

Analysis Variable : DaysOut								
RatingPhrase	N Obs	Mean	Std Dev	Minimum	Maximum	Mode	N	Median
Didn't like it	4	9.0000000	3.3665016	7.0000000	14.0000000	7.0000000	4	7.5000000
Liked it	27	8.7777778	4.2275046	3.0000000	17.0000000	6.0000000	27	7.0000000
Really liked it	73	9.6301370	5.5515435	3.0000000	27.0000000	7.0000000	73	8.0000000
Loved it	25	11.1600000	9.9149718	4.0000000	49.0000000	7.0000000	25	7.0000000

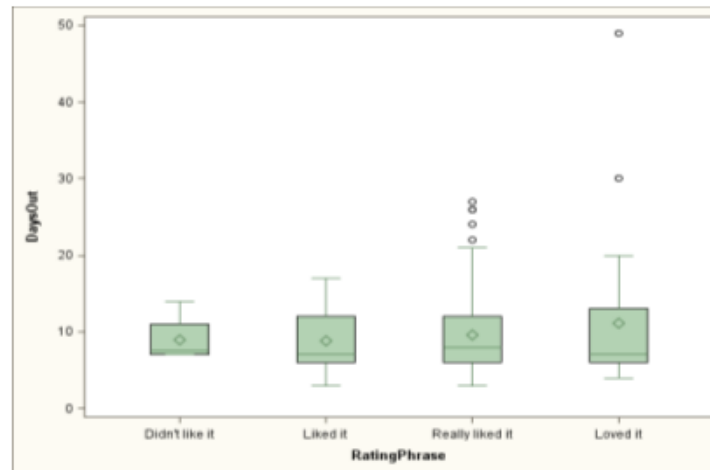
Did we like these movies?

How many?

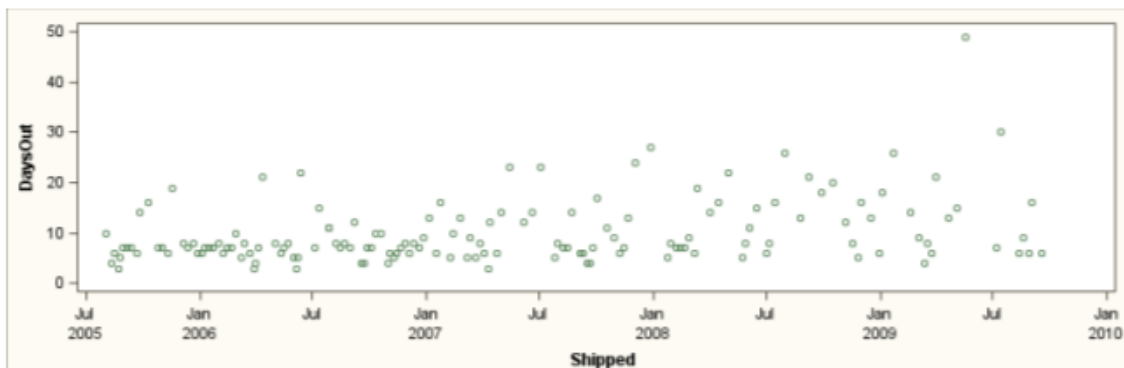


Summary Statistics

Box and Whisker Plots



Days out vs. When shipped



Movie tastes: the extremes

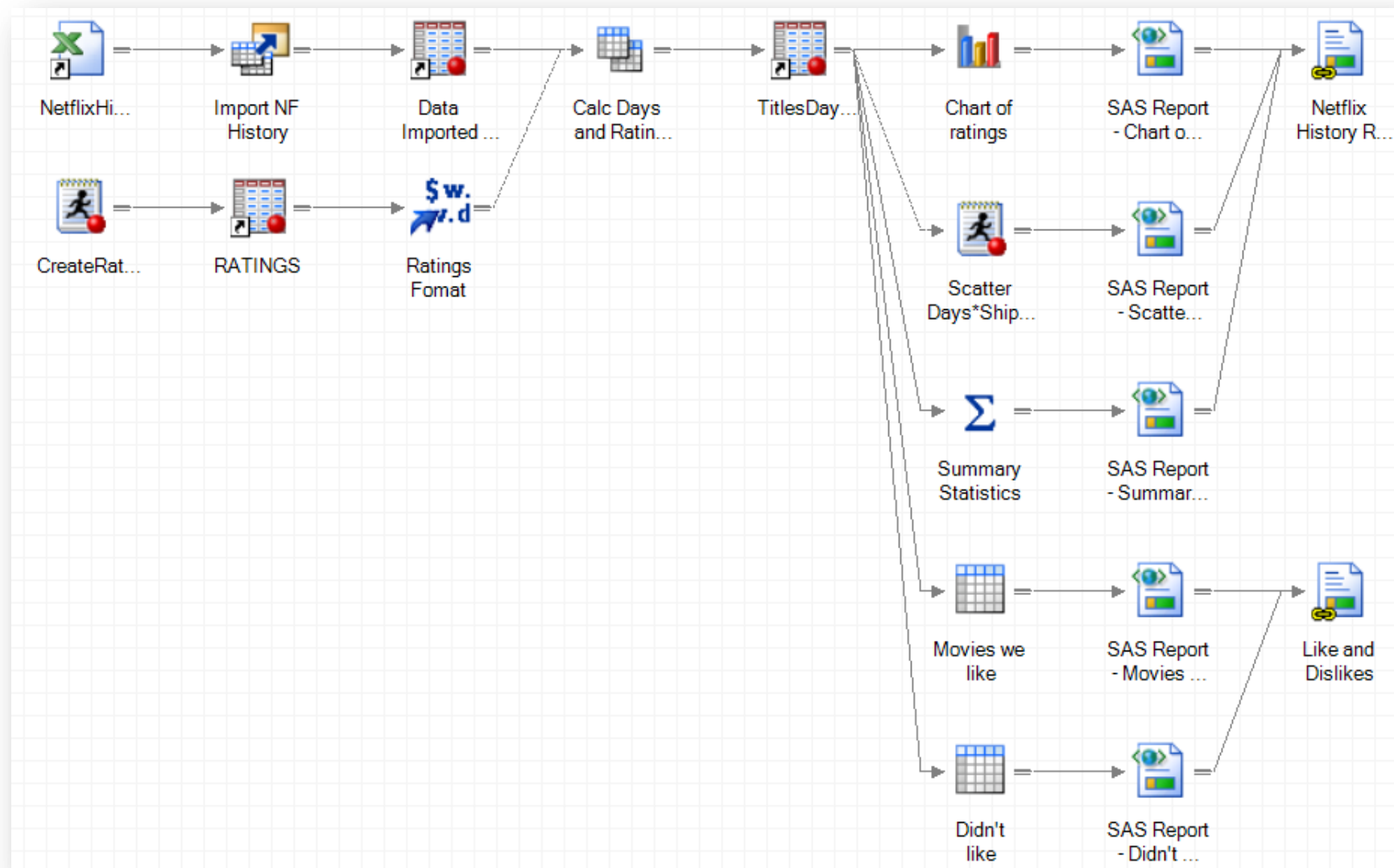
Movies we loved

Movie title	Date watched
Abbott & Costello: Vol. 1: Disc 2	05DEC2005
Best in Show	11SEP2007
Best of Abbott & Costello: Vol. 3: Disc 4	25MAR2009
Bottle Rocket	21SEP2007
Charlotte's Web	03SEP2009
Dreamgirls	30DEC2008
Evan Almighty	28MAY2008
Everything Is Illuminated	13AUG2009
Flushed Away	25SEP2007
Garden State	30JAN2006
High Fidelity	09NOV2005
Iron Man	05JAN2009
John Adams: Disc 3	26NOV2008
Little Miss Sunshine	19JUN2007
Meet the Robinsons	13FEB2008
Night at the Museum	15MAY2007
Ratatouille	04MAY2009
Sideways	06SEP2005
The Curious Case of Benjamin Button	14JUL2009
The Darjeeling Limited	06NOV2008
The Jungle Book	08NOV2007
The Madness of King George	07JUL2009
The Simpsons Movie	29JAN2008
The Young Ones: Disc 1	13DEC2005
Walk the Line	05JUN2006

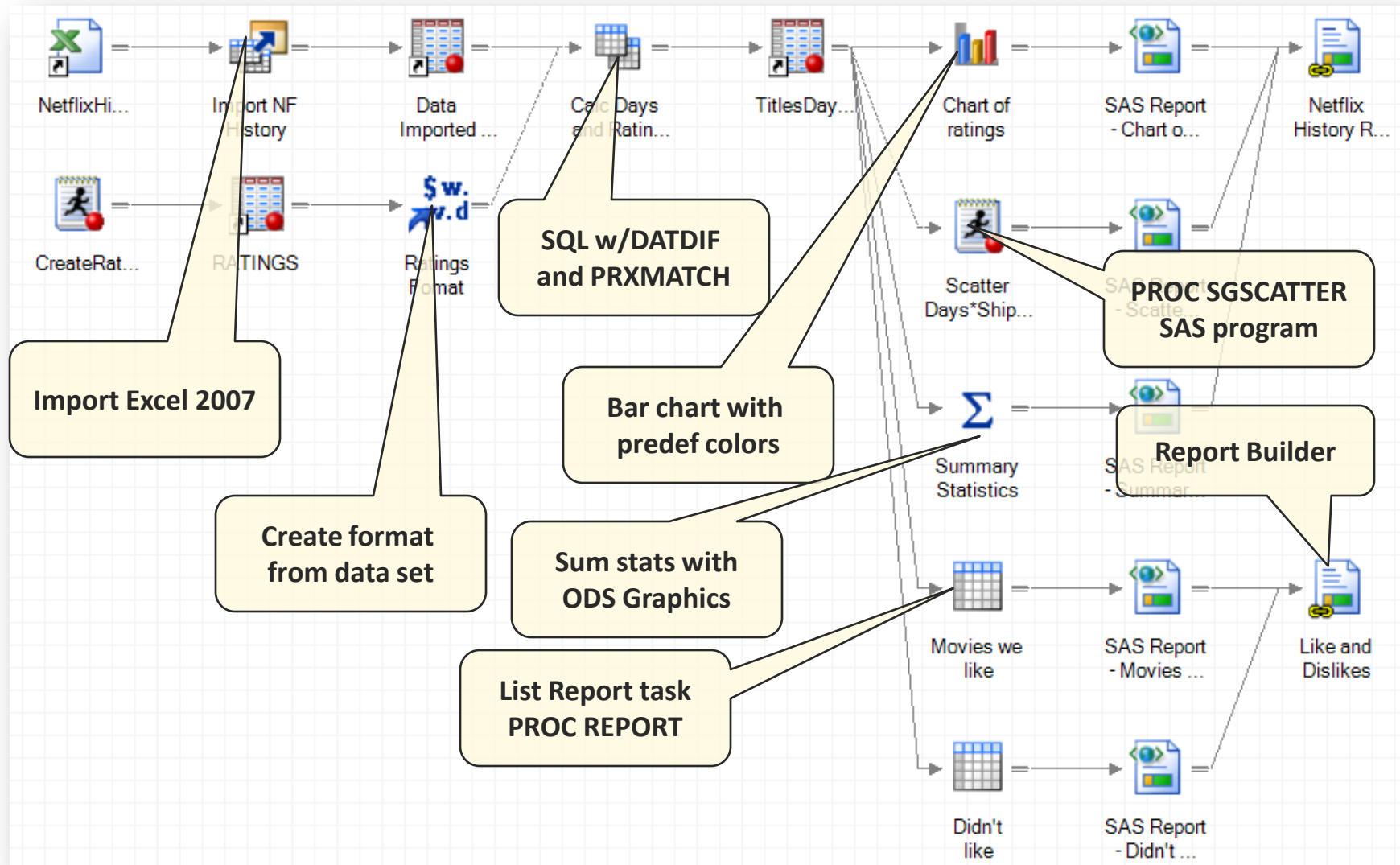
Movies we didn't like

Movie title	Date watched
Be Cool	02NOV2005
Jumper	04MAR2009
Spanglish	31MAY2006
What the #\$*! Do We Know!?	16JAN2006

How it's done



How it's done – features revealed



Analyze movie rental history

DEMO

Screenshot: Import Excel data

Import Data from NetflixHistory.xlsx

3 of 4 Define Field Attributes

Select columns and define attributes:

Inc	Source Name	Name	Label	Type	
<input checked="" type="checkbox"/>	DVD Title	DVD Title	DVD Title	String	\$0
<input checked="" type="checkbox"/>	Rating	Rating	Rating	String	\$0
<input checked="" type="checkbox"/>	Shipped	Shipped	Shipped	Date	D
<input checked="" type="checkbox"/>	Returned	Returned	Returned	Date	A
<input checked="" type="checkbox"/>	Details	Details	Details	String	\$0

Select All Clear All

<Back Next>

Data Imported from NetflixHistory.xlsx

Filter and Sort Query Builder Data Describe Graph Analyze Export Send To

	DVD Title	Rating	Shipped	Returned	Details
1	Rocket Science	You rated this m...	25SEP2009		Report Problem
2		Rate 5 stars			
3		Rate 4 stars			
4		Rate 3 stars			
5		Rate 2 stars			
6		Rate 1 star			
7		Rate not interest...			
8	Before the Devil...	You rated this m...	19SEP2009	25SEP2009	
9		Rate 5 stars			
10		Rate 4 stars			
11		Rate 3 stars			
12		Rate 2 stars			
13		Rate 1 star			
14		Rate not interest...			
15	Children of Heav...	You rated this m...	03SEP2009	19SEP2009	
16		Rate 5 stars			
17		Rate 4 stars			
18		Rate 3 stars			
19		Rate 2 stars			
20		Rate 1 star			
21		Rate not interest...			
22	Charlotte's Web	You rated this m...	28AUG2009	03SEP2009	
23		Rate 5 stars			
24		Rate 4 stars			
25		Rate 3 stars			
26		Rate 2 stars			
27		Rate 1 star			
28		Rate not interest...			
29	Indiana Jones/Ki...	You rated this m...	19AUG2009	28AUG2009	

Screenshot: Filter and calculate

Calc Days and Ratings for Local:WORK.NETFLIXHISTORY

Query name: Output name: [Change...](#)

Computed Columns | Prompt Manager | Preview | Tools | Options

[Add Tables](#) [Delete](#) [Join Tables](#)

t1 (NETFLIXHISTORY)

- DVD Title
- Rating
- Shipped
- Returned
- Details
- Computed Columns
 - DaysOut
 - ActualRating
 - RatingPhrase

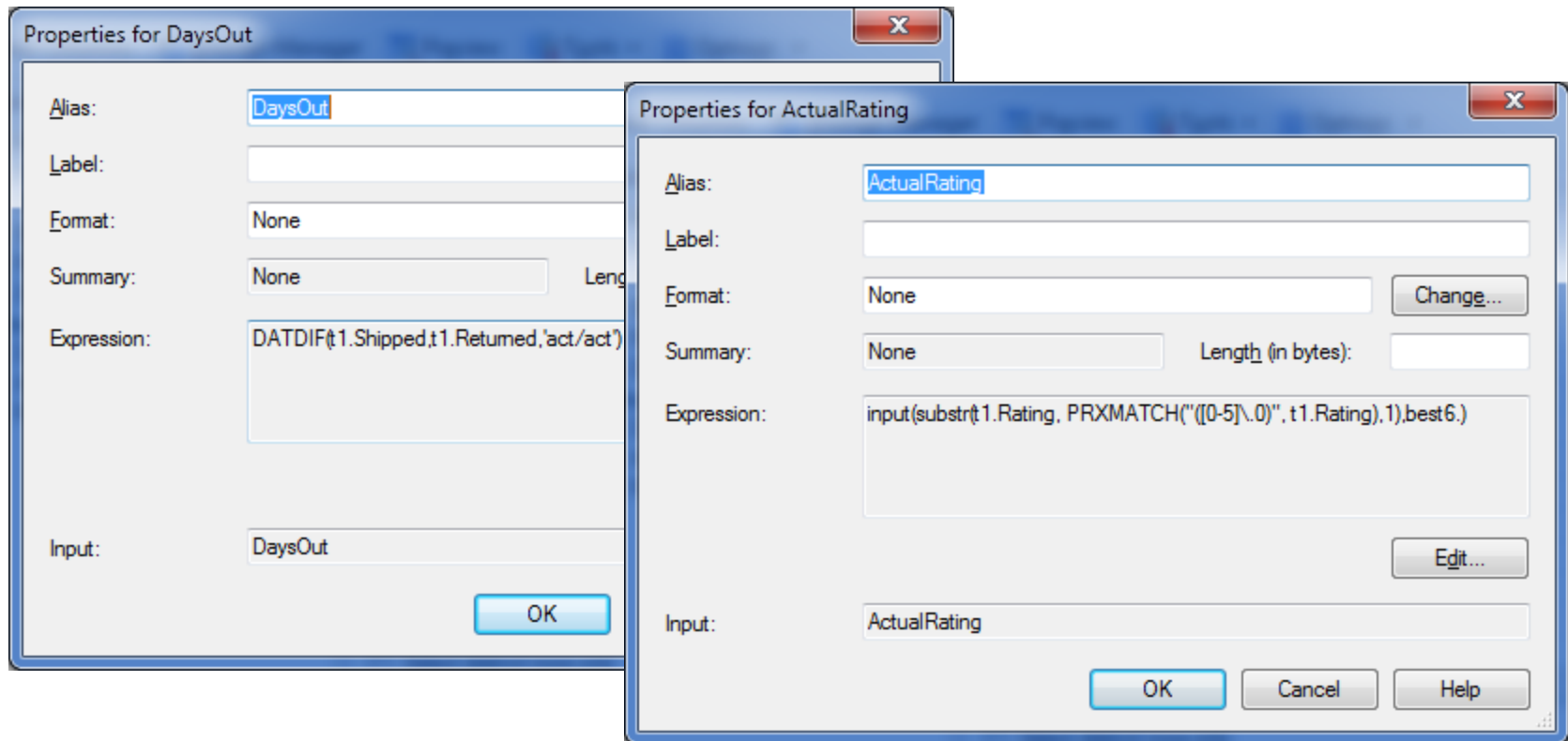
Select Data | Filter Data | Sort Data

Column Name	Input	Summary
DVD Title	t1.DVD Title	
Shipped	t1.Shipped	
Returned	t1.Returned	
DaysOut	DaysOut	
ActualRating	ActualRating	
RatingPhrase	RatingPhrase	

☐ Select distinct rows only

[Run](#) [Save and Close](#) [Cancel](#) [Help](#)

Screenshot: calculations



`input(substr(t1.Rating, PRXMATCH("[0-5]\.0)", t1.Rating),1),best6.)`

Calc Days and Ratings ▾

Input Data Code Log Output Data

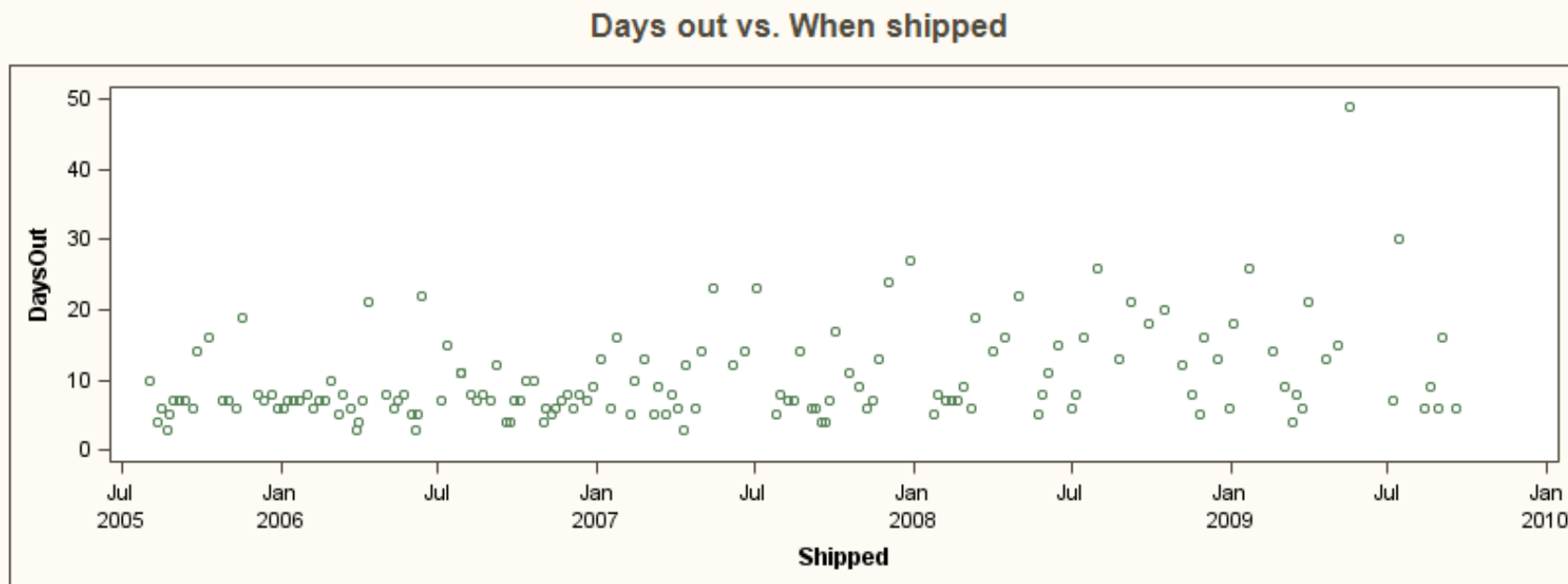
Modify Task Filter and Sort Query Builder Data ▾ Describe ▾ Graph ▾ Analyze ▾ Export ▾ Send

	DVD Title	Shipped	Returned	DaysOut	ActualRating	RatingPhrase
1	Before the Devil...	19SEP2009	25SEP2009	6	3	Liked it
2	Children of Heav...	03SEP2009	19SEP2009	16	4	Really liked it
3	Charlotte's Web	28AUG2009	03SEP2009	6	5	Loved it
4	Indiana Jones/Ki...	19AUG2009	28AUG2009	9	4	Really liked it
5	Babel	13AUG2009	19AUG2009	6	4	Really liked it
6	Everything Is Illu...	14JUL2009	13AUG2009	30	5	Loved it
7	The Curious Cas...	07JUL2009	14JUL2009	7	5	Loved it
8	The Madness of...	19MAY2009	07JUL2009	49	5	Loved it
9	Sweeney Todd	04MAY2009	19MAY2009	15	4	Really liked it
10	Ratatouille	21APR2009	04MAY2009	13	5	Loved it
11	Mrs. Brown	31MAR2009	21APR2009	21	4	Really liked it
12	Casino Royale	25MAR2009	31MAR2009	6	3	Liked it
13	Best of Abbott &...	17MAR2009	25MAR2009	8	5	Loved it
14	The Dark Knight	13MAR2009	17MAR2009	4	4	Really liked it
15	Monsters, Inc.	04MAR2009	13MAR2009	9	4	Really liked it

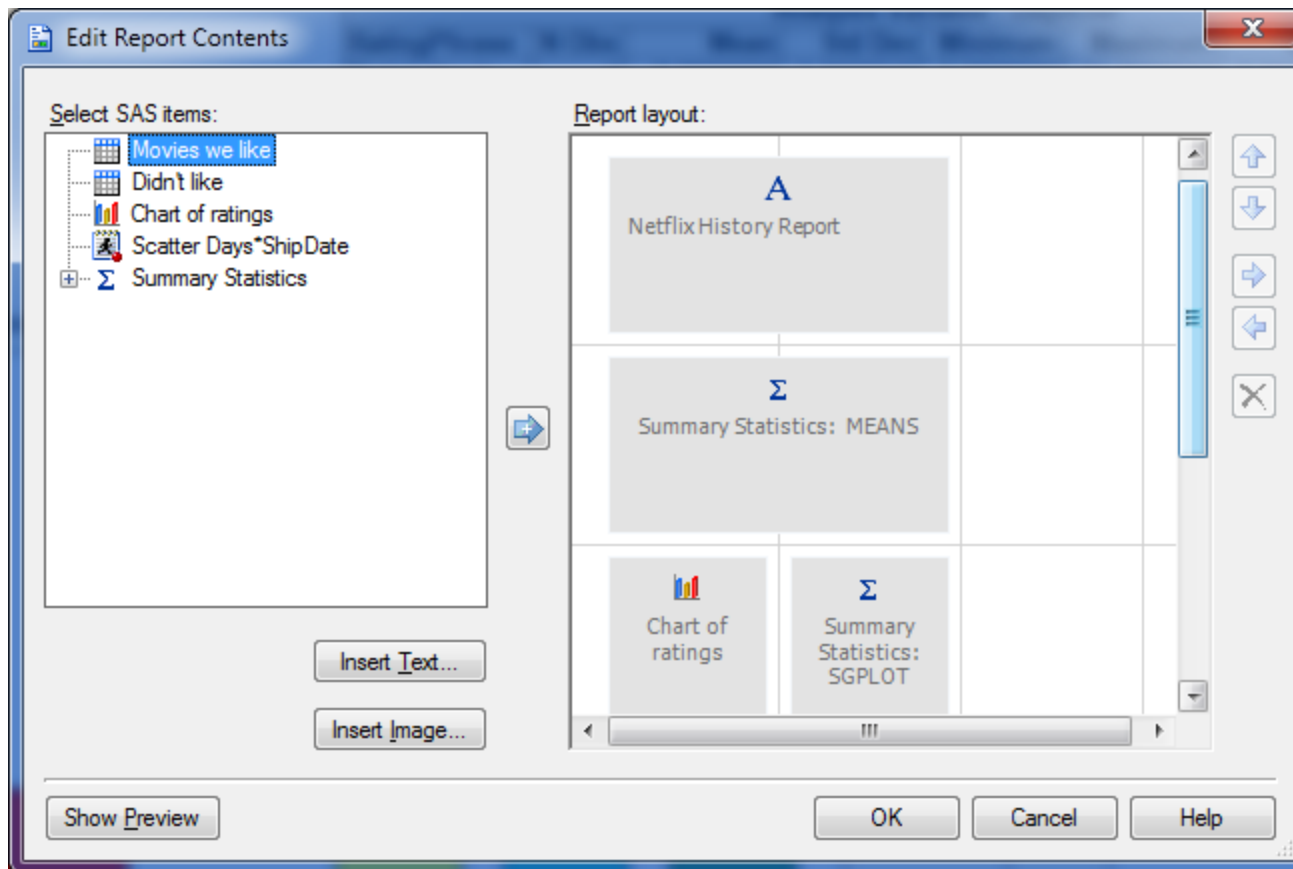
Screenshot: SGCATTER

```
ods graphics / width=1200 height=400 imagemap=on;  
title "Days out vs. When shipped";
```

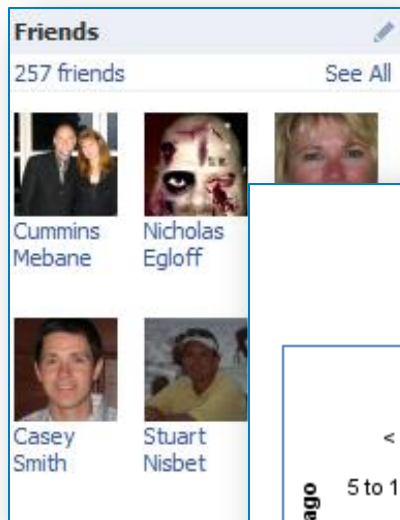
```
proc sgscatter data=work.TITLESDBSRATINGS;  
  plot DaysOut*Shipped;  
run;  
ods graphics off;
```



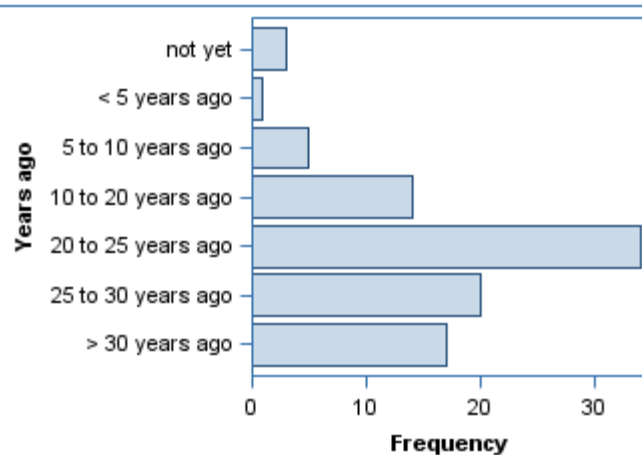
Screenshot: Report builder



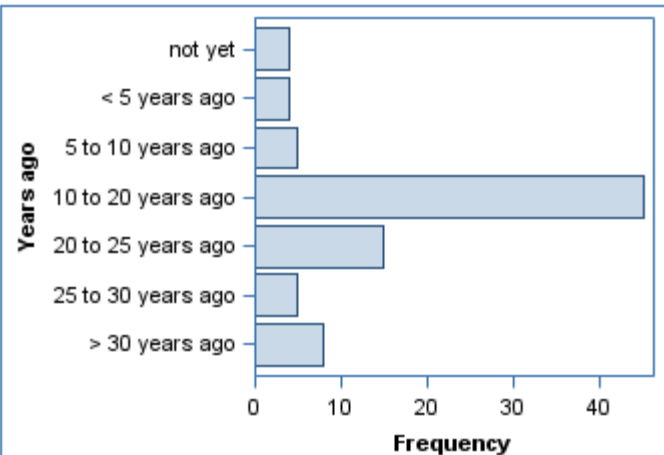
Facebook: Who are my friends, really?



When my friends graduated high school



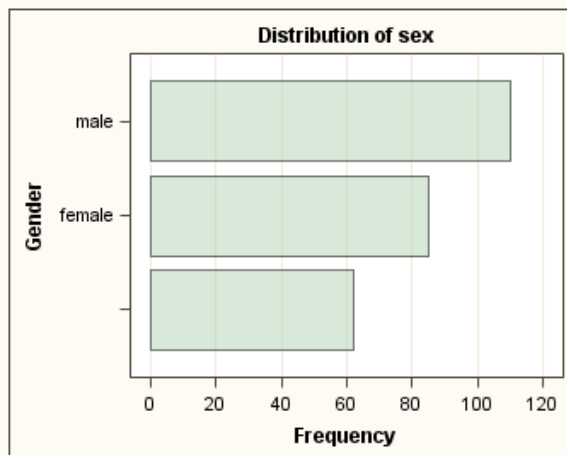
When my friends graduated college



Friend count: Gender

The FREQ Procedure

Gender				
sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	62	24.12	62	24.12
female	85	33.07	147	57.20
male	110	42.80	257	100.00



Friend count: Country

The FREQ Procedure

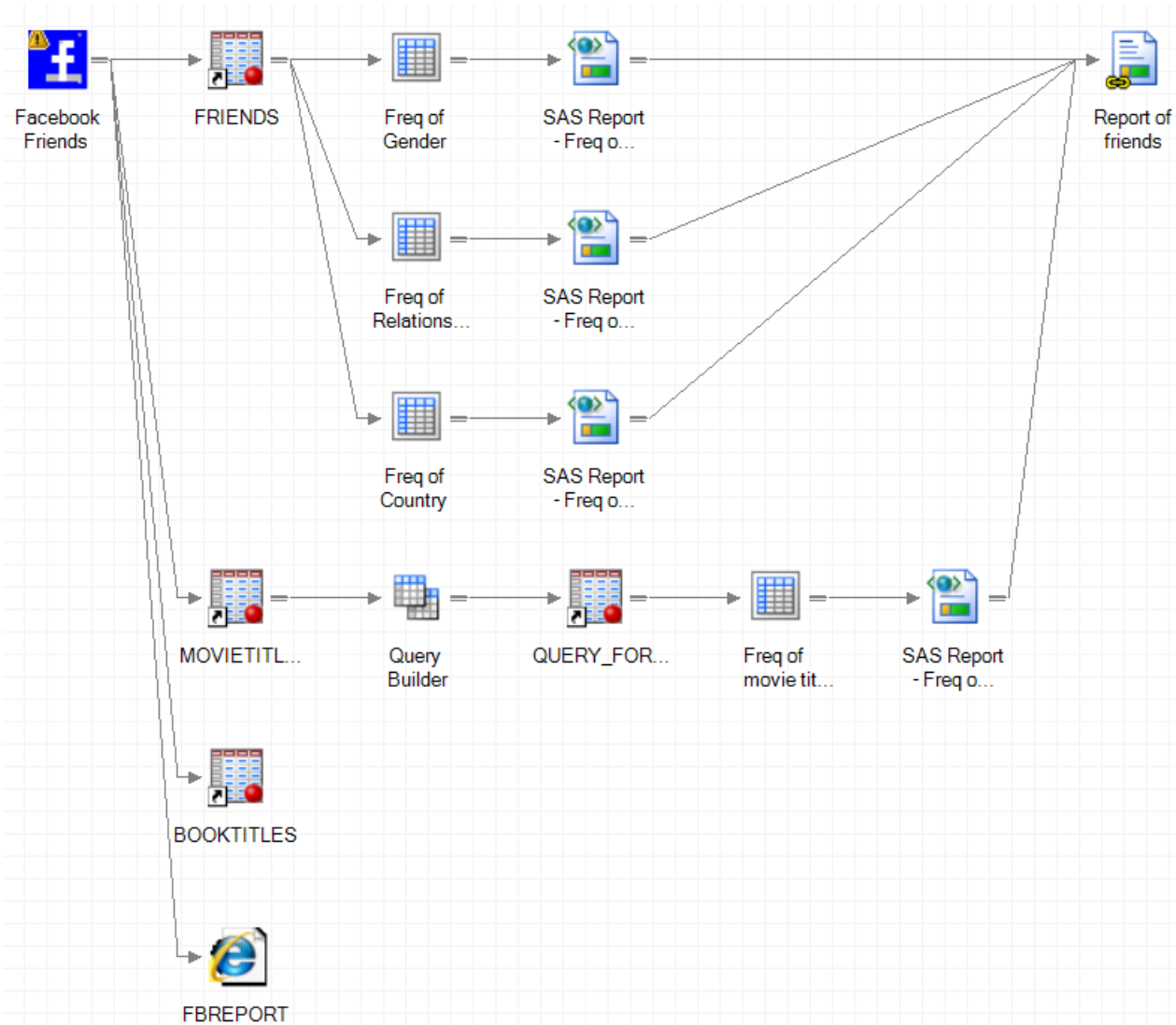
country				
country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	162	63.04	162	63.04
Canada	5	1.95	167	64.98
Hong Kong	1	0.39	168	65.37
Singapore	1	0.39	169	65.76
United States	88	34.24	257	100.00

Some favorite movies

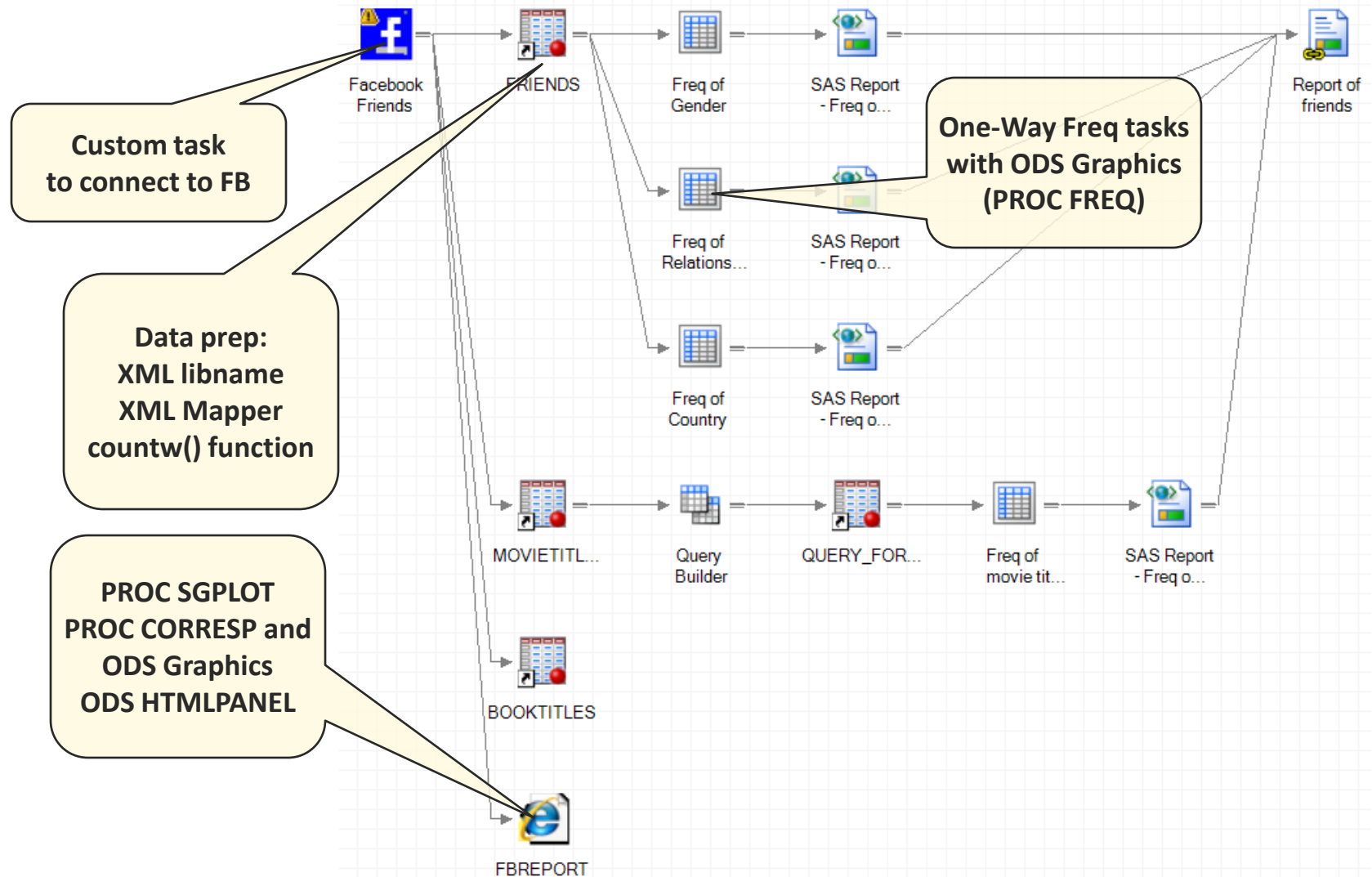
The FREQ Procedure

normalizedtitle	Frequency
star wars	10
shawshank redemption	8
indiana jones	5
animal house	4
lord of the rings	4
back to the future	3
caddyshack	3
casablanca	3
dead poets society	3
etc.	3
ferris bueller's day off	3
fight club	3
forrest gump	3
the shawshank redemption	3
a few good men	2
apollo 13	2
batman: the dark knight	2
braveheart	2
dirty harry	2
evil dead 2	2
field of dreams	2
good will hunting	2
grease	2
napoleon dynamite	2
once	2
sound of music	2
superbad	2
the usual suspects	2
top gun	2
tremors	2
"the wall" "rocky horror pictu	1
...etc.	1
12	1
13	1
2001 - a space odyssey monty p	1

How it's done



How it's done – features revealed!



Analyze Facebook friends

DEMO

We ♥ SAS programmers

what's coming in SAS Enterprise Guide 4.3

- Automatic syntax suggestions and completion
- Integrated syntax help
- “Tidy” feature for SAS and macro programs
- Code analyzer to view your program flow
- And flying cars

Syntax suggestions and completion

```

/* summarize the data across a category and store */
/* the output in an output data set */
proc means data=&data &stat noprint;
    var &measure;
    class &category &report / mi;
    output out=summary &stat=&
run;

/* store the value of the meas
/* the row count into a macro
/* later in the report */
proc sql noprint;
select &measure, _FREQ_ into :o
from summary where _TYPE_=0;
select count(distinct &category) into :ca
quit;

```

- ASCEND
- ASCENDING
- DESCEND
- DESCENDING
- EXCLUSIVE
- GROUPINTERNAL
- MISSING
- MLF
- ORDER=

```

/* summarize the data across a category and store */
/* the output in an output data set */
proc means data=&data &stat noprint;
    var &
    class
    output
run;

/* store
/* the ro
/* later
proc sql

```

- category
- categorycount
- data
- measure
- measureformat
- n
- numobs
- overall
- report

```

&measure &category /leve
measure for ALL rows and
variable for use */

```

Integrated syntax help

```

/* summarize the data across a category and store */
/* the output in an output data set */
proc means data=&data &stat noprint;
run;
/* st
/* th
/* la
proc
selec
from
selec
quit;

/* sort the results so that we get the TOP values */
/* rising to the top of the data set */

```

MEANS

Context: [PROCEDURE DEFINITION] PROC MEANS

Syntax: PROC MEANS <option(s)> <statistic-keyword(s)>;
BY <DESCENDING> variable-1 <... <DESCENDING> variable-n><NOTSORTED>;
CLASS variable(s) </ option(s)>;
FREQ variable;
ID variable(s);
OUTPUT <OUT=SAS-data-set> <output-statistic-specification(s)> <id-group-specification(s)> <maximum-id-specification(s)>
<minimum-id-specification(s)> </ option(s)> ;
TYPES request(s);
VAR variable(s) < / WEIGHT=weight-variable>;
WAYS list;
WEIGHT variable;

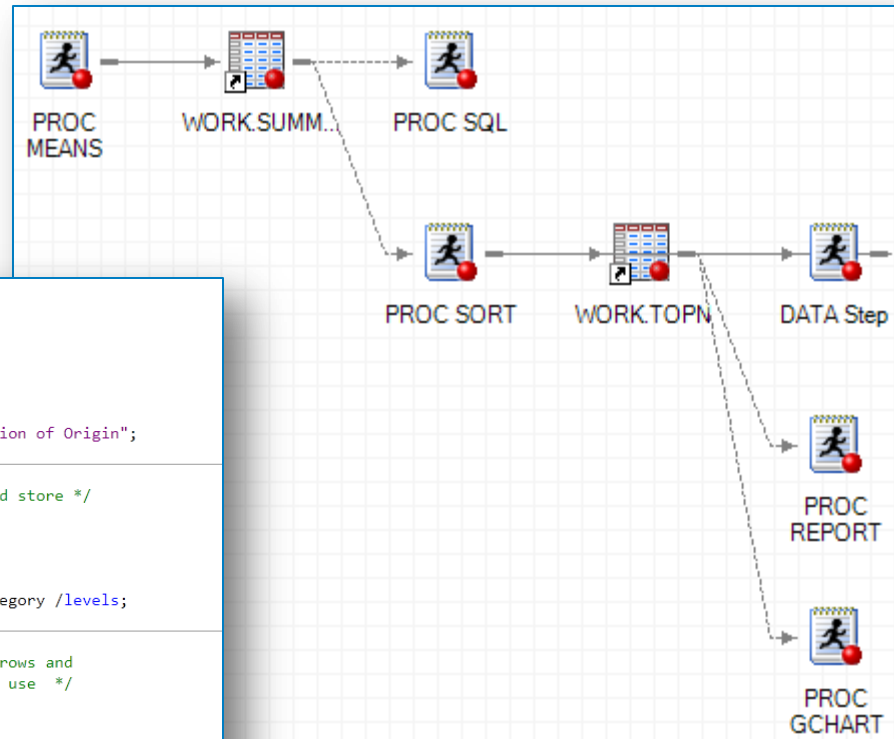
The MEANS procedure provides data summarization tools to compute descriptive statistics for variables across all observations and within groups of observations.

“Tidy” up your programs

```
/* Pass through the data and output the first N */  
/* values for each category */  
data topn;  
  length rank 8; label rank="Rank"; set topn;  
  by &category descending &measure; if first.&category then rank=0;  
  rank+1; if rank le &n then output;  
run;
```

```
/* Pass through the data and output the first N */  
/* values for each category */  
data topn;  
  length rank 8;  
  label rank="Rank";  
  
  set topn;  
  by &category descending &measure;  
  
  if first.&category then  
    rank=0;  
  rank+1;  
  
  if rank le &n then  
    output;  
run;
```

Analyze your SAS program flow



```

%let data=SASHELP.CARS;
%let report=Model;
%let measure=MPG_City;
%let measureformat=%str(format=BEST6.);
%let stat=MEAN;
%let n=10;
%let category=Origin;
title "Top Models by MPG_City for each region of Origin";
footnote;

/* summarize the data across a category and store */
/* the output in an output data set */
proc means data=&data &stat noprint;
  var &measure;
  class &category &report;
  output out=summary &stat=&measure &category /levels;
run;

/* store the value of the measure for ALL rows and
/* the row count into a macro variable for use */
/* later in the report */
proc sql noprint;
  select &measure, _FREQ_ into :overall, :numobs
  from summary where _TYPE_=0;
  select count(distinct &category) into :categorycount from summary;
quit;

/* sort the results so that we get the TOP values */
/* rising to the top of the data set */
proc sort data=work.summary out=work.topn;
  where _type_>2;
  by &category descending &measure;
run;
  
```


More information

- <http://support.sas.com/eguide>
- <http://blogs.sas.com/sasdummy>
- Support.sas.com: Discussion forums
- SAS Global Forum papers: search for “Enterprise Guide”
- On Twitter? Follow me [@cjdinger](#) (and find lots of others as well)



**THE
POWER
TO KNOW®**