A Practical Example of SGPLOT Using Logistic Regression

Jon Yankey

Clinical Trials and Statistical Data Management Center

Department of Biostatistics

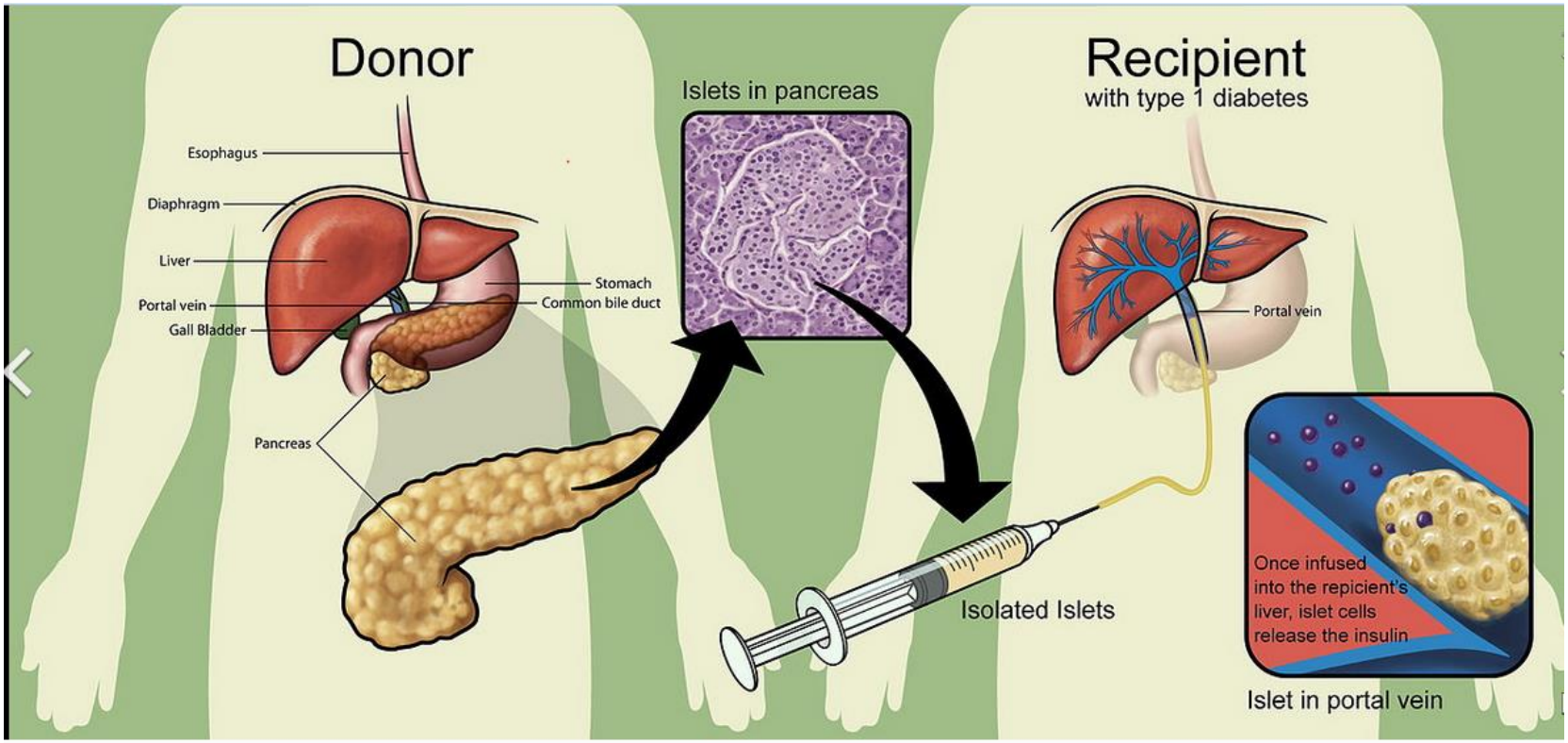
University of Iowa

Background

Clinical Islet Transplantation Consortium

- **Clinical Islet Transplantation (CIT) Consortium established by NIAID and NIDDK in 2004**
- **Goals are to:**
  - **Advance islet transplantation through innovative Phase 2 trials and to**
  - **Obtain licensure for an islet product through Phase 3 trials for use in Type 1 diabetes with severe hypoglycemia**
- **9 Clinical Sites in US and Canada, 3 Clinical Sites in Sweden and Norway**
- **Clinical Trials Statistical and Data Management Center, University of Iowa, serves as the statistical and data coordinating center**

Background



Islet transplantation PLoS Medicine CC BY 2.5 view terms

Giovanni Maki - Naftanel MA, Harlan DM (2004) Pancreatic Islet Transplantation. PLoS Med 1(3): e58 (image link)

Background

Clinical investigators wanted to know if rejection of transplanted islet cells was related to the number of HLA (Human Leukocyte Antigen) anitbody mis-matches

Antibodies are produced by both the patient and the transplanted islet cells

If patient does not have antibodies similar to the transplanted cells:

- **Patient's immune system recognizes transplanted cells as intruders**
- **Immune system response is to remove transplanted cells**
- **Transplant did not work – transplanted cells rejected**

Investigators were interested in A/B and D/R antibody mis-matches

Background

Goal of analyses:

- **Use logistic regression to estimate odds of transplant rejection as function of number of mis-matches for A/B and D/R antibodies**
- **Summarize results graphically**

Background

Contrived data from sashelp.cars data set

Outcome is MPG greater than or equal to 25 (*Islet cells rejected or not*)

Predictor variable is overall vehicle length (*Number of HLA mismatches*)

Fit two separate models:
one for Rear wheel drive vehicles *(Number of A/B mismatches)* &
one for Front wheel drive vehicles *(Number of D/R mismatches)*

Two Separate Logistic regression models:

$\text{Logit}(\text{MPG} \geq 25) = \beta_0 + \beta_1 * (\text{Vehicle length})$ *Rear wheel only*

$\text{Logit}(\text{MPG} \geq 25) = \beta_0 + \beta_2 * (\text{Vehicle length})$ *Front wheel only*

Background

The MEANS Procedure

Analysis Variable : MPG_Highway MPG (Highway)

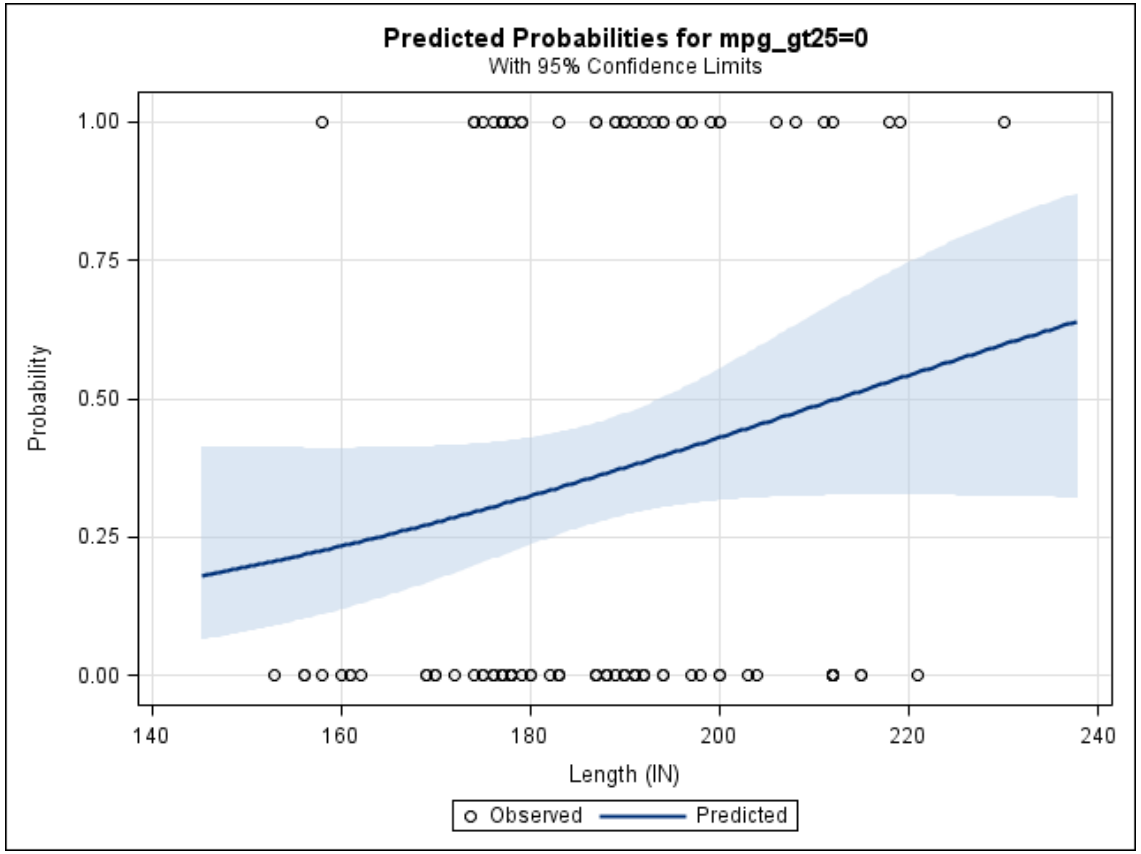
Drive Train	N Obs	N	Mean	Std Dev	Minimum	Maximum
All	92	92	22.47	4.16	12.00	29.00
Front	226	226	29.50	5.89	18.00	66.00
Rear	110	110	25.04	3.00	18.00	36.00

Table of DriveTrain by mpg_gt25

DriveTrain mpg_gt25

Frequency Col Pct	0	1	Total
All	56 45.16	36 11.84	92
Front	28 22.58	198 65.13	226
Rear	40 32.26	70 23.03	110
Total	124	304	428

Graph Produced by PROC Logistic

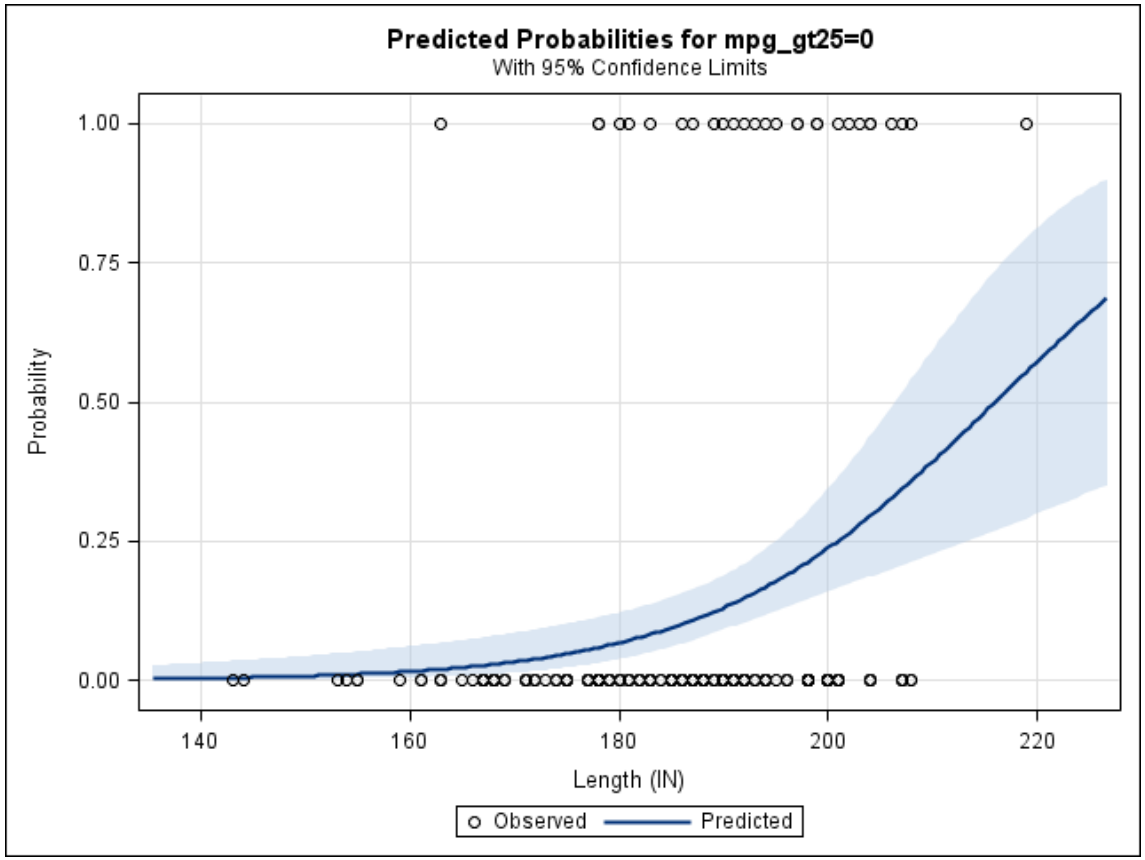


Estimated odds-ratio of MPG ge 25 per 1 inch increase in vehicle length:

1.02

95% CI (0.99, 1.05)
(Rear wheel drive only)

Graph Produced by PROC Logistic

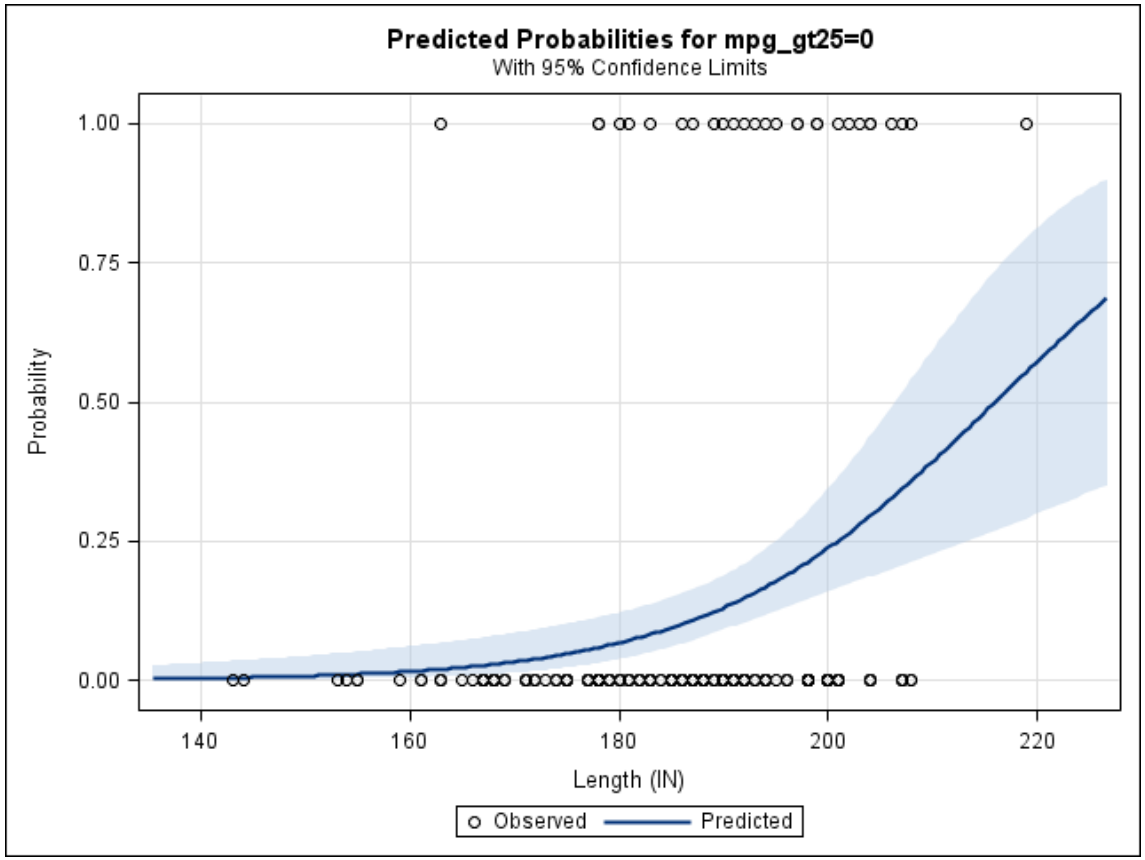


Estimated odds-ratio of MPG ge 25 per 1 inch increase in vehicle length:

1.07

95% CI (1.03, 1.12)
(Front wheel drive only)

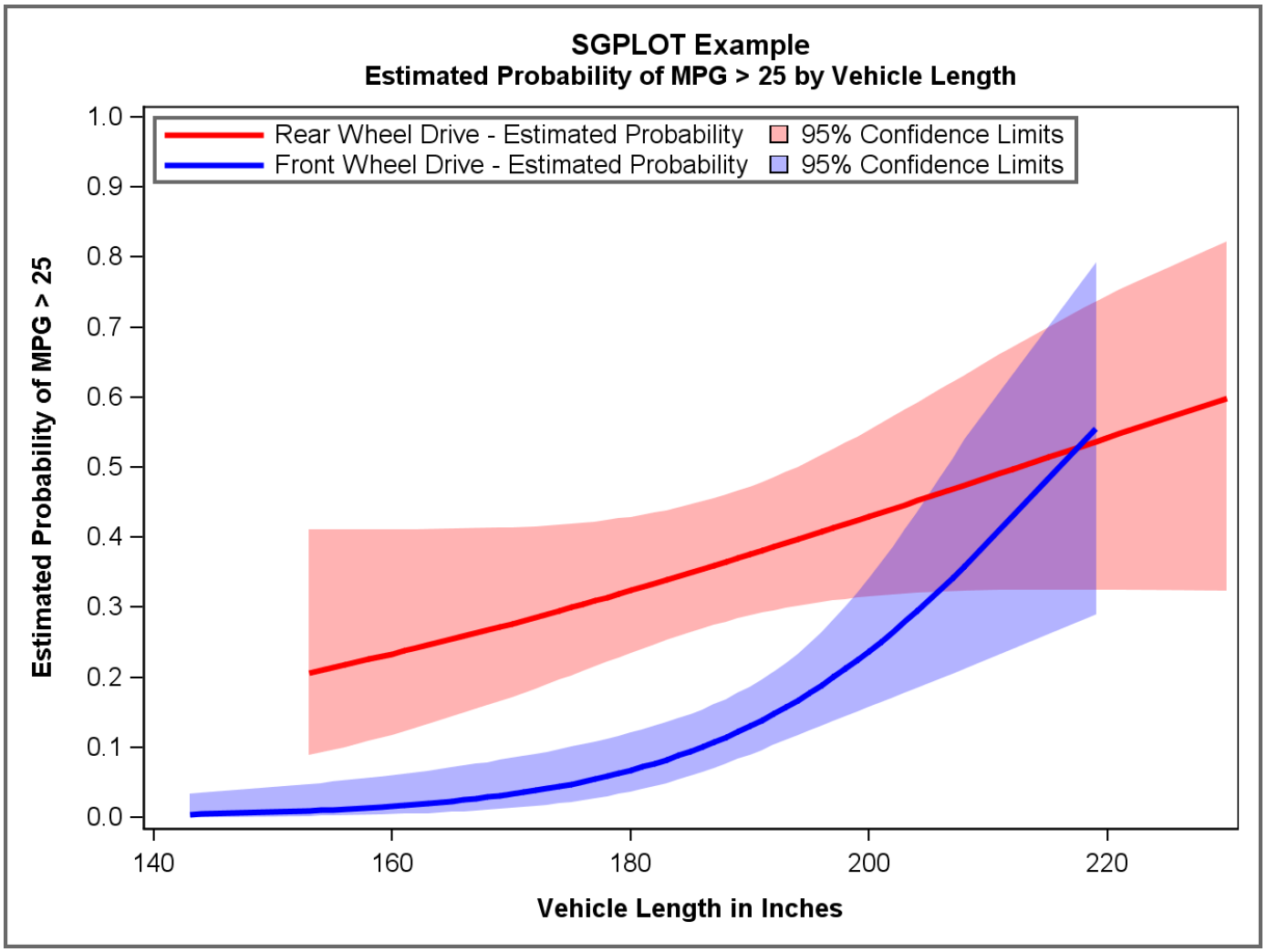
Graph Produced by PROC Logistic



Nice graphs produced but wanted to customize:

- **Title**
- **Axes labels**
- **Remove the open circle 'observed' data points.**
- **Remove grid lines**
- **Overlay both plots onto a single graph**

Customized Graph



Customization of Graph

To customize graphs:

- **Need estimated probabilities for rear and front wheel drive vehicles**
- **Need lower 95% CI for estimated probabilities that correspond to each observed length rear wheel drive used in model**
- **Need upper 95% CI for estimated probabilities that correspond to each observed length rear wheel drive used in model**
- **Explore options in SG procedures for different plot types , options for titles, and options for x and y-axes.**

Obtaining Estimated Probabilities

Statements used to fit logistic regression models:

```
proc logistic data = cars plots=all;
        model mpg_gt25 = length;
        where drivetrain = 'Rear';
        output  out = rear
                p = p_rear
                xbeta = X_rear
                lower = lower_rear
                upper = upper_rear;
run;
```

- Restrict observations to rear wheel only
- Create data set that contains:
- Estimated probabilities
- Observed vehicle length (or linear predictor)
- Lower 95% CI for given vehicle length
- Upper 95% CI for given vehicle length

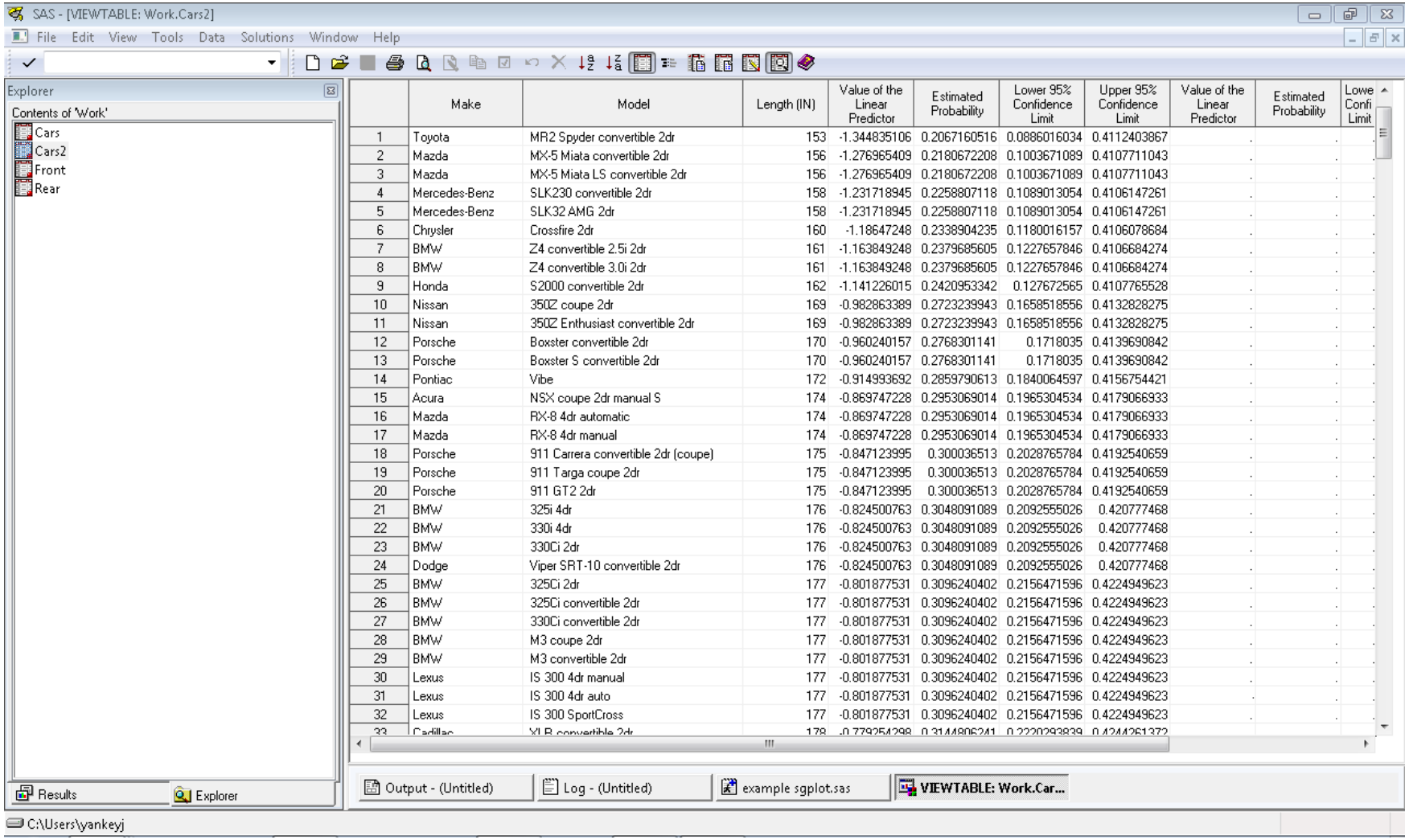
For Front wheel drive model, change where statement to 'where drivtrain = 'Front'; ' and change name of output data set 'output out = front … '

Data Management

Some data management:

```
proc sort data = rear
        (keep = make model p_rear x_rear lower_rear upper_rear length);
        by length;
run;

proc sort data = front
        (keep = make model p_front x_front lower_front upper_front length);
        by length;
run;

data cars2;
        set rear
        front;
run;
```

Data Management



	Make	Model	Length (IN)	Value of the Linear Predictor	Estimated Probability	Lower 95% Confidence Limit	Upper 95% Confidence Limit	Value of the Linear Predictor	Estimated Probability	Lowe Confi Limit
1	Toyota	MR2 Spyder convertible 2dr	153	-1.344835106	0.2067160516	0.0886016034	0.4112403867	.	.	.
2	Mazda	MX-5 Miata convertible 2dr	156	-1.276965409	0.2180672208	0.1003671089	0.4107711043	.	.	.
3	Mazda	MX-5 Miata LS convertible 2dr	156	-1.276965409	0.2180672208	0.1003671089	0.4107711043	.	.	.
4	Mercedes-Benz	SLK230 convertible 2dr	158	-1.231718945	0.2258807118	0.1089013054	0.4106147261	.	.	.
5	Mercedes-Benz	SLK32 AMG 2dr	158	-1.231718945	0.2258807118	0.1089013054	0.4106147261	.	.	.
6	Chrysler	Crossfire 2dr	160	-1.18647248	0.2338904235	0.1180016157	0.4106078684	.	.	.
7	BMW	Z4 convertible 2.5i 2dr	161	-1.163849248	0.2379685605	0.1227657846	0.4106684274	.	.	.
8	BMW	Z4 convertible 3.0i 2dr	161	-1.163849248	0.2379685605	0.1227657846	0.4106684274	.	.	.
9	Honda	S2000 convertible 2dr	162	-1.141226015	0.2420953342	0.127672565	0.4107765528	.	.	.
10	Nissan	350Z coupe 2dr	169	-0.982863389	0.2723239943	0.1658518556	0.4132828275	.	.	.
11	Nissan	350Z Enthusiast convertible 2dr	169	-0.982863389	0.2723239943	0.1658518556	0.4132828275	.	.	.
12	Porsche	Boxster convertible 2dr	170	-0.960240157	0.2768301141	0.1718035	0.4139690842	.	.	.
13	Porsche	Boxster S convertible 2dr	170	-0.960240157	0.2768301141	0.1718035	0.4139690842	.	.	.
14	Pontiac	Vibe	172	-0.914993692	0.2859790613	0.1840064597	0.4156754421	.	.	.
15	Acura	NSX coupe 2dr manual S	174	-0.869747228	0.2953069014	0.1965304534	0.4179066933	.	.	.
16	Mazda	RX-8 4dr automatic	174	-0.869747228	0.2953069014	0.1965304534	0.4179066933	.	.	.
17	Mazda	RX-8 4dr manual	174	-0.869747228	0.2953069014	0.1965304534	0.4179066933	.	.	.
18	Porsche	911 Carrera convertible 2dr (coupe)	175	-0.847123995	0.300036513	0.2028765784	0.4192540659	.	.	.
19	Porsche	911 Targa coupe 2dr	175	-0.847123995	0.300036513	0.2028765784	0.4192540659	.	.	.
20	Porsche	911 GT2 2dr	175	-0.847123995	0.300036513	0.2028765784	0.4192540659	.	.	.
21	BMW	325i 4dr	176	-0.824500763	0.3048091089	0.2092555026	0.420777468	.	.	.
22	BMW	330i 4dr	176	-0.824500763	0.3048091089	0.2092555026	0.420777468	.	.	.
23	BMW	330Ci 2dr	176	-0.824500763	0.3048091089	0.2092555026	0.420777468	.	.	.
24	Dodge	Viper SRT-10 convertible 2dr	176	-0.824500763	0.3048091089	0.2092555026	0.420777468	.	.	.
25	BMW	325Ci 2dr	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
26	BMW	325Ci convertible 2dr	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
27	BMW	330Ci convertible 2dr	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
28	BMW	M3 coupe 2dr	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
29	BMW	M3 convertible 2dr	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
30	Lexus	IS 300 4dr manual	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
31	Lexus	IS 300 4dr auto	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
32	Lexus	IS 300 SportCross	177	-0.801877531	0.3096240402	0.2156471596	0.4224949623	.	.	.
33	Cadillac	XLR convertible 2dr	178	-0.779254298	0.3144806241	0.2220293839	0.4244261372			

Full Code for Customized Graph

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25 by Vehicle Length';

        band x = length upper= upper_rear lower=lower_rear /  fillattrs=(color=red)
                transparency = 0.70 name='rear band' legendlabel='95% Confidence Limits';

        band x = length upper= upper_front lower=lower_front /
                fillattrs=(color=blue) transparency = 0.70 name='front band' legendlabel='95% Confidence Limits';

        series x = length y = p_rear /
                lineattrs=(color=red thickness = 3) name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';

        series x = length y = p_front /
                lineattrs=(color=blue thickness = 3)
                name = 'pred front' legendlabel = 'Front Wheel Drive - Estimated Probability';

        keylegend 'pred rear' 'rear band' 'pred front' 'front band' / across = 2 location = inside position = topleft;

        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';

        xaxis label = 'Vehicle Length in Inches';
run;
```

Full Code for Customized Graph

Steps towards full code

- **Searched SAS references (books and papers) for examples**

Step 1:

- **Confidence band for estimated probabilities over observed range of vehicle lengths for rear wheel drive vehicles only**
- **Titles, labeling of axes, legend, etc.**

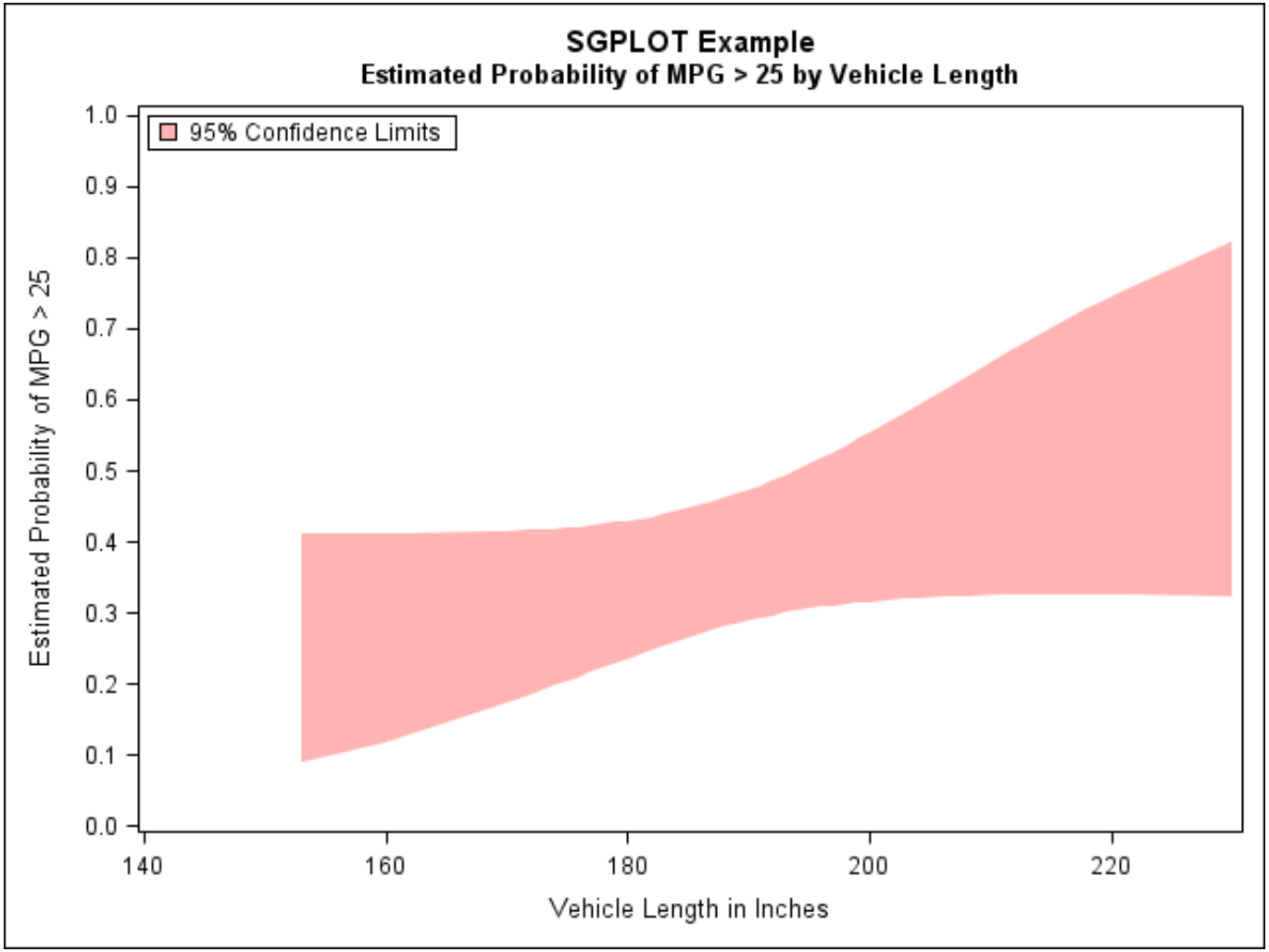
Step 2:

- **Line plot for estimated probabilities**
- **Update legend**

Step 3:

- **Use lessons learned from Steps 1 and 2 to add confidence band and line plot for estimated probabilities for front wheel drive vehicles**

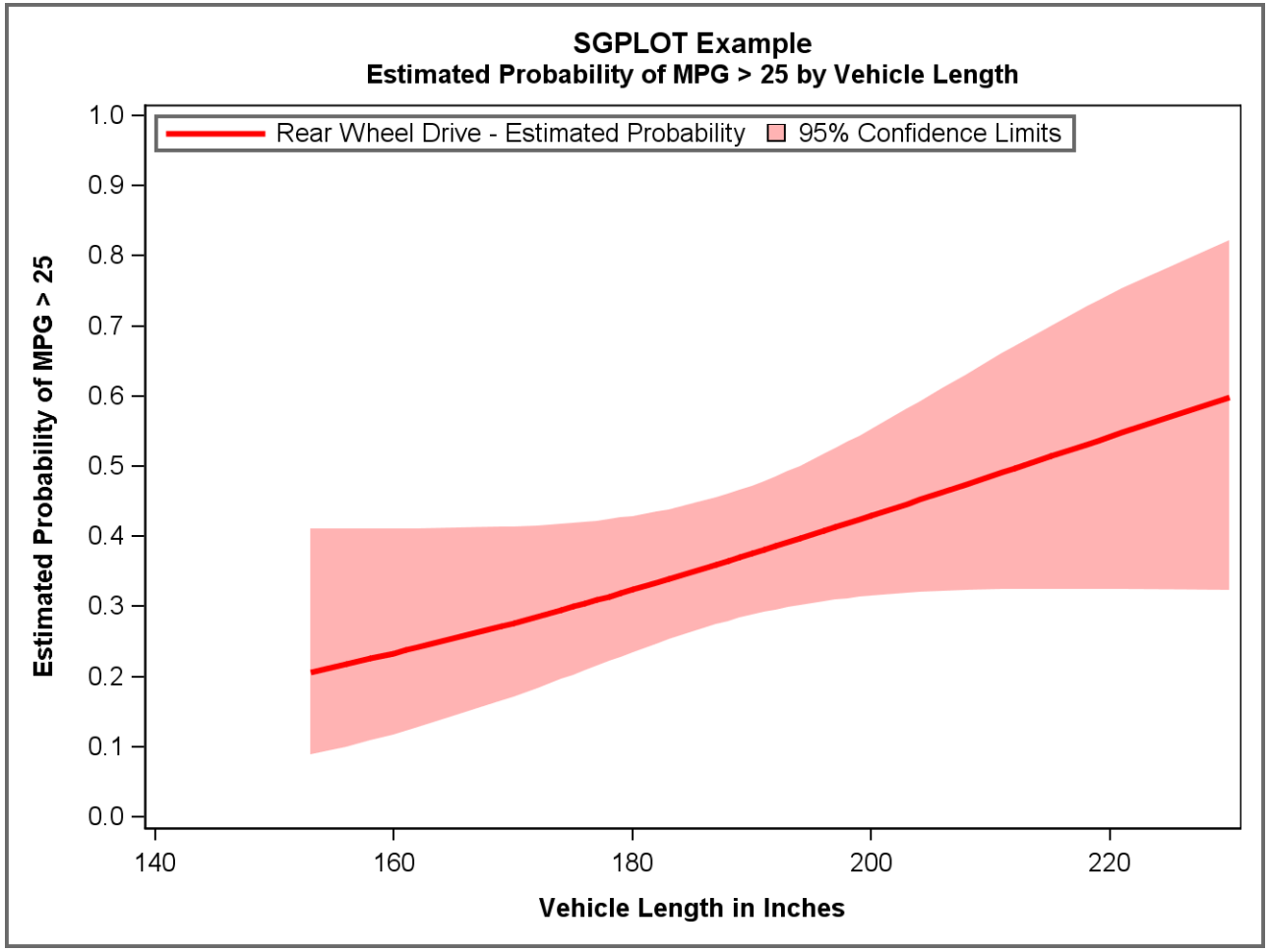
Step 1


SGPLOT Example
Estimated Probability of MPG > 25 by Vehicle Length
95% Confidence Limits
Estimated Probability of MPG > 25
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
140
160
180
200
220
Vehicle Length in Inches

Step 1

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 ‘Estimated Probability of MPG > 25’;

        band x = length upper= upper_rear lower=lower_rear /
                fillattrs=(color=red) transparency = 0.70
                name='rear band'
                legendlabel='95% Confidence Limits';


        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';

        xaxis label = ‘Vehicle Length in Inches’;
run;
```

Step 2



2nd Step:

Add estimated probabilities for Rear wheel drive and update legend

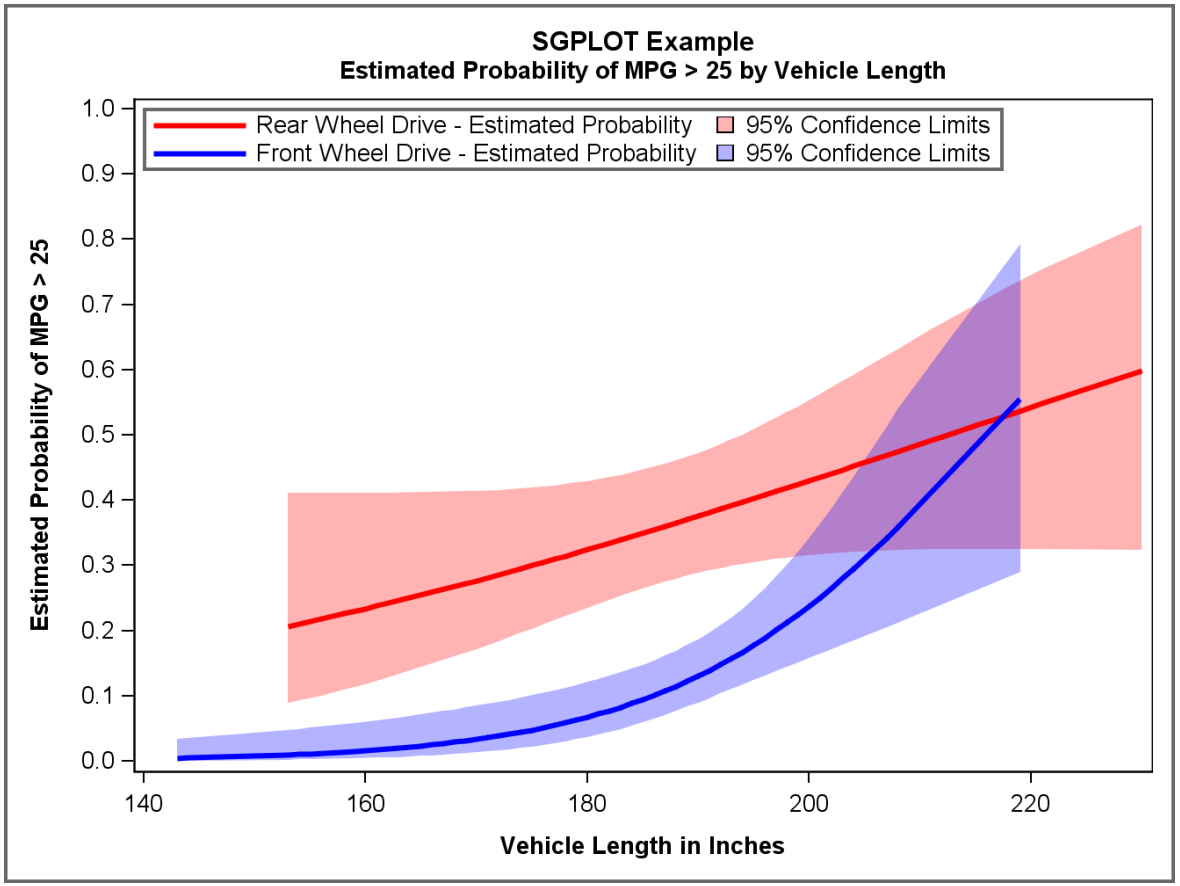
Step 2

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25';

        band x = length upper= upper_rear lower=lower_rear /
                name='rear band'
                legendlabel='95% Confidence Limits';

        series x = length y = p_rear /
                lineattrs=(color=red thickness = 3)
                name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';

        keylegend 'pred rear' 'rear band' / across = 2 location = inside position = topleft;

        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';

        xaxis label = 'Vehicle Length in Inches';
run;
```

Step 1

Step 2

Step 3



3rd step:

Add CI band and estimated probabilities for Front wheel drive and update legend

Step 3

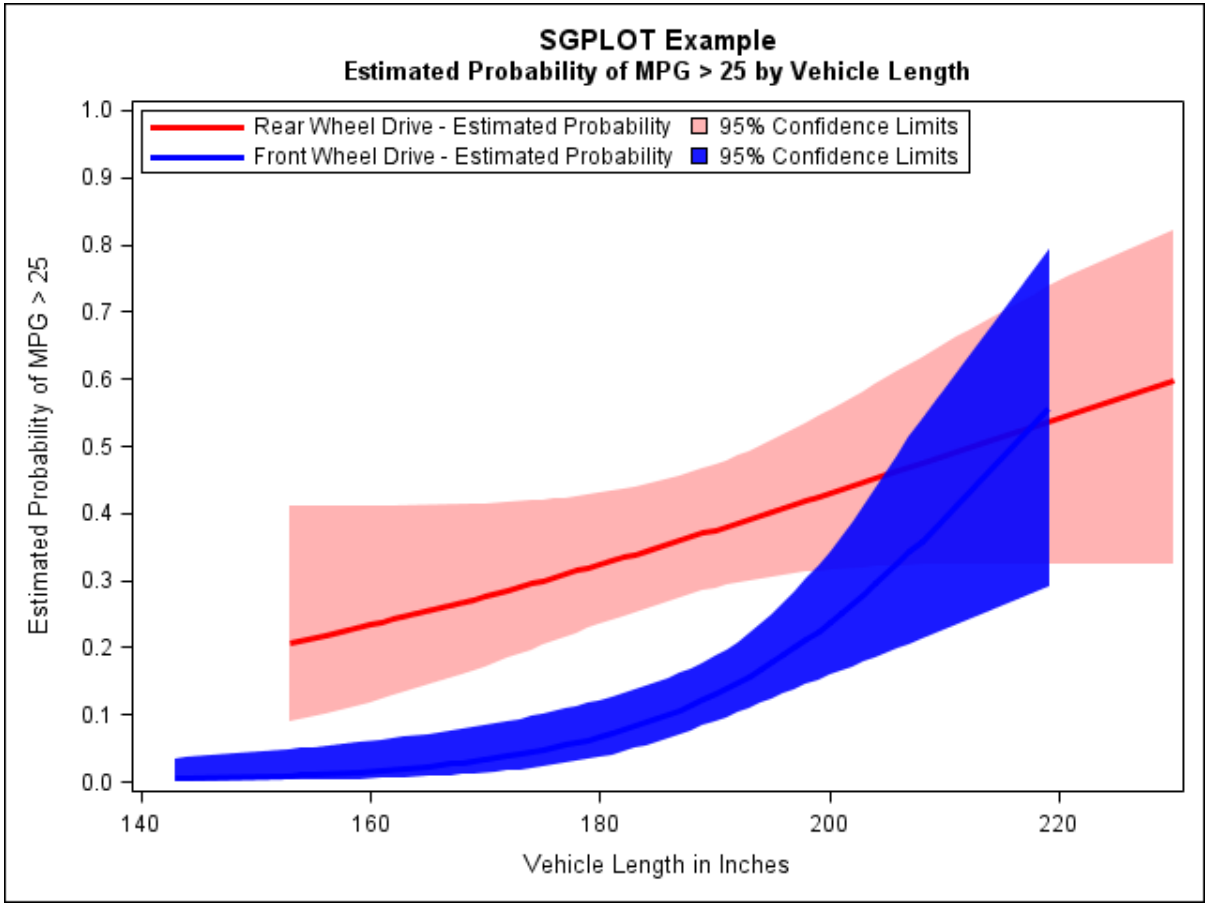
```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25 by Vehicle Length';

Step 1  band x = length upper= upper_rear lower=lower_rear /  fillattrs=(color=red)
                transparency = 0.70 name='rear band' legendlabel='95% Confidence Limits';

Step 3  band x = length upper= upper_front lower=lower_front /
                fillattrs=(color=blue) transparency = 0.70 name='front band' legendlabel='95% Confidence Limits';

Step 2  series x = length y = p_rear /
                lineattrs=(color=red thickness = 3) name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';

Step 3  series x = length y = p_front /
                lineattrs=(color=blue thickness = 3)
                name = 'pred front' legendlabel = 'Front Wheel Drive - Estimated Probability';

Step 3  keylegend 'pred rear' 'rear band' 'pred front' 'front band' / across = 2 location = inside position = topleft;

        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';

        xaxis label = 'Vehicle Length in Inches';
run;
```

Order of Plot Statements

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25 by Vehicle Length';
        band x = length upper= upper_rear lower=lower_rear /  fillattrs=(color=red)
                transparency = 0.70 name='rear band' legendlabel='95% Confidence Limits';

        band x = length upper= upper_front lower=lower_front /
                fillattrs=(color=blue) transparency = 0.70 name='front band' legendlabel='95% Confidence Limits';

        series x = length y = p_rear /
                lineattrs=(color=red thickness = 3) name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';

        series x = length y = p_front /
                lineattrs=(color=blue thickness = 3)
                name = 'pred front' legendlabel = 'Front Wheel Drive - Estimated Probability';

        keylegend 'pred rear' 'rear band' 'pred front' 'front band' / across = 2 location = inside position = topleft;
        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';
        xaxis label = 'Vehicle Length in Inches';
run;
```

Order of Plots Matters

Step 3



95% CI band is blocking out other plots…

Adjust transparency level

Change order of plots

Naming and Labeling of Each Plot

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25 by Vehicle Length';
        band x = length upper= upper_rear lower=lower_rear /  fillattrs=(color=red)
                transparency = 0.70 name='rear band' legendlabel='95% Confidence Limits';

        band x = length upper= upper_front lower=lower_front /
                fillattrs=(color=blue) transparency = 0.70 name='front band' legendlabel='95% Confidence Limits';

        series x = length y = p_rear /
                lineattrs=(color=red thickness = 3) name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';

        series x = length y = p_front /
                lineattrs=(color=blue thickness = 3)
                name = 'pred front' legendlabel = 'Front Wheel Drive - Estimated Probability';

        keylegend 'pred rear' 'rear band' 'pred front' 'front band' / across = 2 location = inside position = topleft;
        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';
        xaxis label = 'Vehicle Length in Inches';
run;
```

Order of Plots Matters

Customizing Legend

```
proc sgplot data = cars2;
        title 'SGPLOT Example';
        title2 'Estimated Probability of MPG > 25 by Vehicle Length';
        band x = length upper= upper_rear lower=lower_rear /  fillattrs=(color=red)
                transparency = 0.70 name='rear band' legendlabel='95% Confidence Limits';   1

        band x = length upper= upper_front lower=lower_front /
                fillattrs=(color=blue) transparency = 0.70 name='front band' legendlabel='95% Confidence Limits';
                                                                                            2
        series x = length y = p_rear /
                lineattrs=(color=red thickness = 3) name = 'pred rear'
                legendlabel = 'Rear Wheel Drive - Estimated Probability';   3

        series x = length y = p_front /
                lineattrs=(color=blue thickness = 3)
                name = 'pred front' legendlabel = 'Front Wheel Drive - Estimated Probability';   4

                3       1       4       2
        keylegend 'pred rear' 'rear band' 'pred front' 'front band' / across = 2 location = inside position = topleft;
        yaxis min = 0 max = 1 values = (0 to 1 by 0.1) label = 'Estimated Probability of MPG > 25';
        xaxis label = 'Vehicle Length in Inches';
run;
```

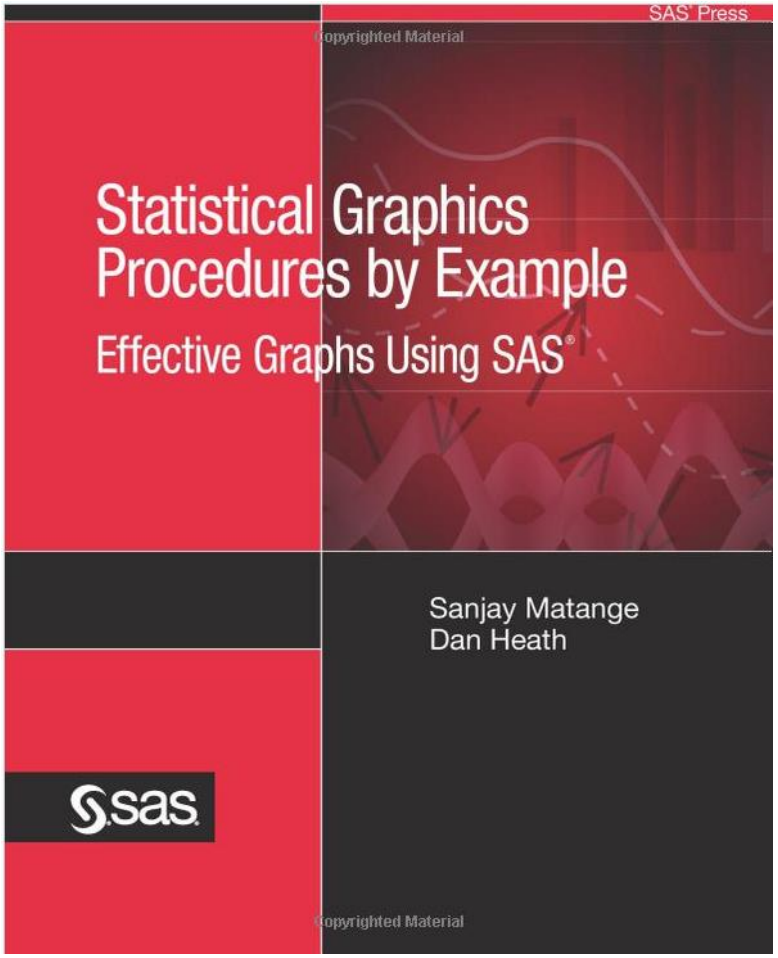
Order of Plots Matters

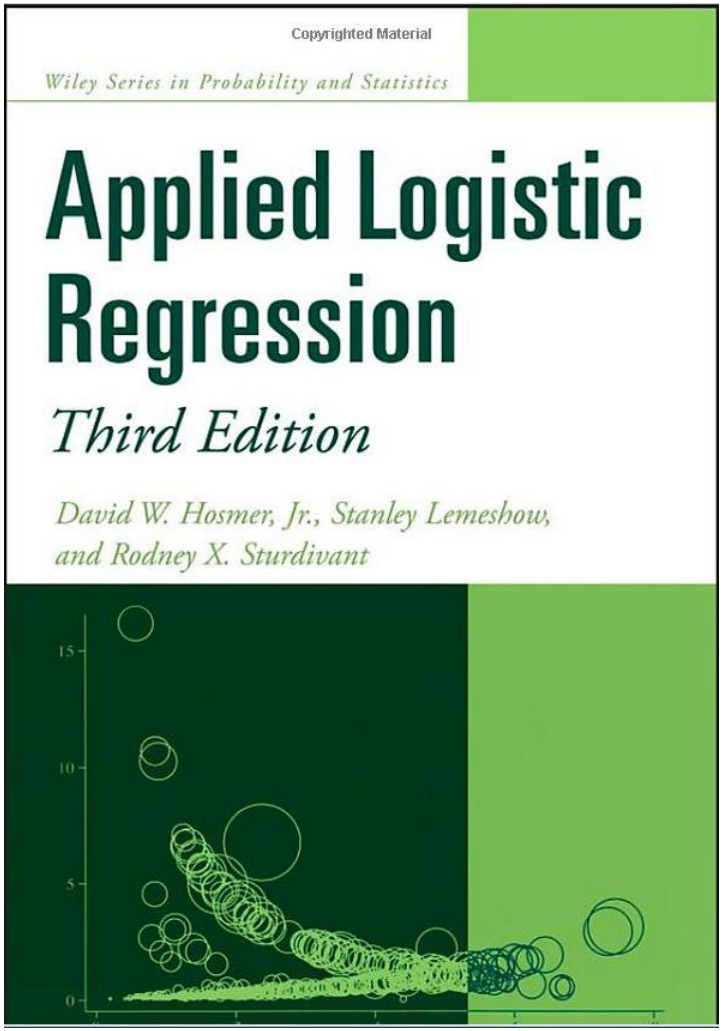
Summary

- **Order of plots matters –**
  - **plots are layered in the order in which they appear after the SGPLOT statement**
  - **1st plot is drawn, the next plot is drawn on top of the first, etc.**
  - **Keep this in mind.**
  - **Can use options for different plot types to alter appearance of plots**

- **Name each plot**
  - **Naming each plot allows user to customize graph legend**
  - **Appearance of legend is not affected by order of the plots**

References


SAS Press
Copyrighted Material
Statistical Graphics Procedures by Example
Effective Graphs Using SAS®
Sanjay Matange
Dan Heath
sas
Copyrighted Material


Copyrighted Material
Wiley Series in Probability and Statistics
Applied Logistic Regression
Third Edition
David W. Hosmer, Jr., Stanley Lemeshow, and Rodney X. Sturdivant