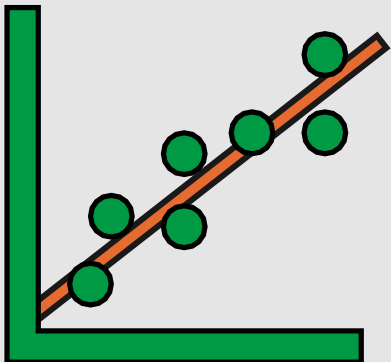
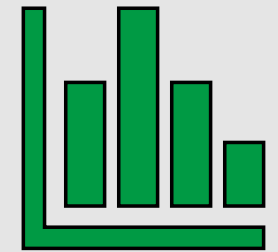


# Making Magnificently Good Graphs: SAS

BERDC Special Topics Talk 3, Part 3



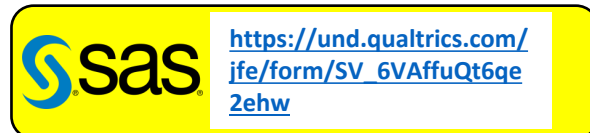
**DaCCoTA**  
DAKOTA CANCER COLLABORATIVE  
ON TRANSLATIONAL ACTIVITY

Dr. Mark Williamson  
Biostatistics, Epidemiology,  
and Research Design Core

# Overview

- We'll cover how to make great looking graphs in SAS
- We'll start by creating basic graphs, then explore how to upgrade by modifying various elements

- Take the pre-test here



- Get the SAS-code here



- Get the SAS dataset here



- Get the PDF version here



- Stay tuned for a neat treat at the end

## Elements:

- I. **Labels**
- II. **Axes**
- III. **Colors and Shapes**
- IV. **Dots, Lines, and Text**
- V. **Other**



# Getting Set Up

## SAS Studio

- Can get set up below:

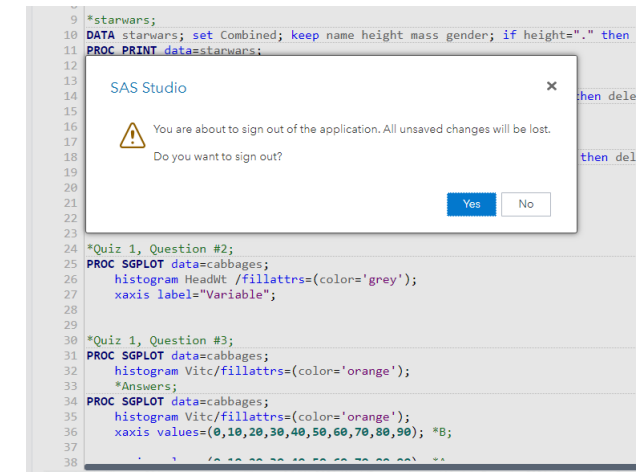
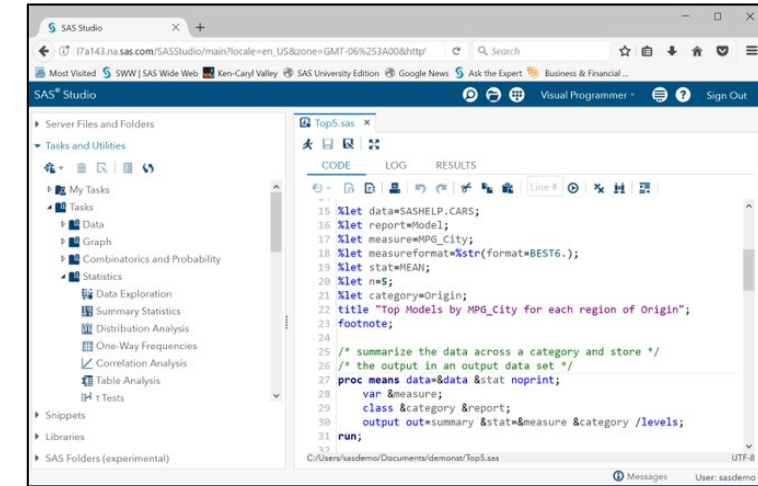
[https://www.sas.com/en\\_us/software/on-demand-for-academics.html#features](https://www.sas.com/en_us/software/on-demand-for-academics.html#features)

- Pros:

- Free
- Great support
- Sign in from any computer

- Cons:

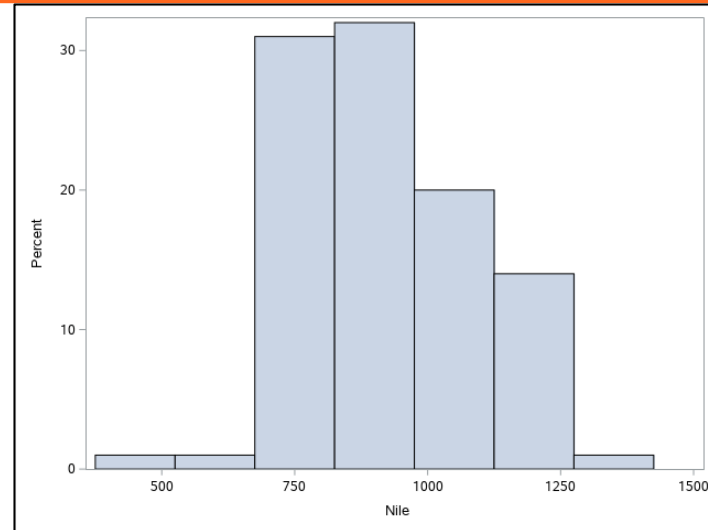
- Not all SAS functionality available in Studio
- Upload limits
- Unsaved work loss



# Histograms

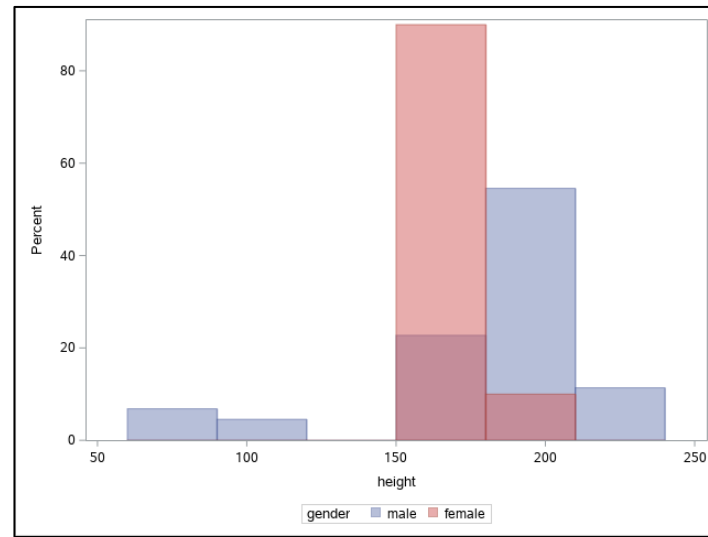
## I. Simple histogram

```
PROC SGPLOT data=Nile;  
    histogram Nile;
```



## II. Two-way histogram

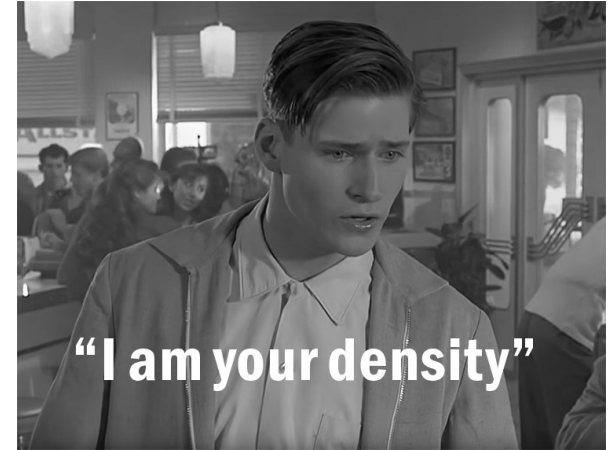
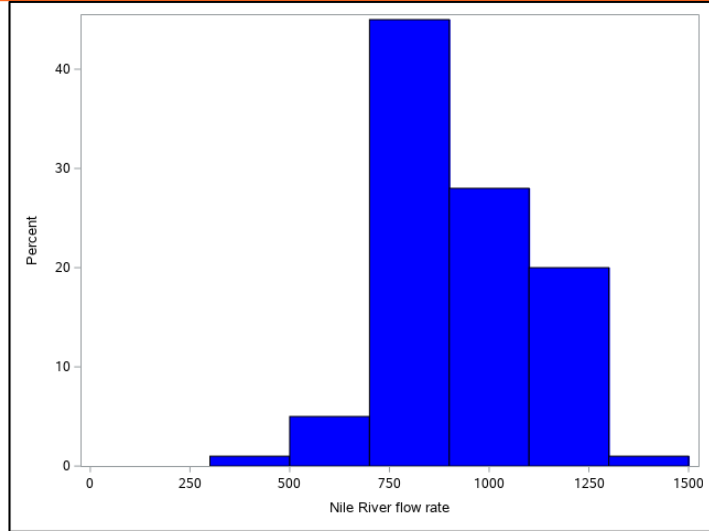
```
PROC SGPLOT data=starwars;  
    histogram height/ group=gender transparency=0.5;
```



# Histograms cont.

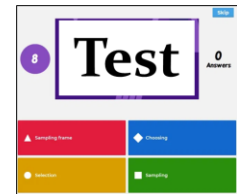
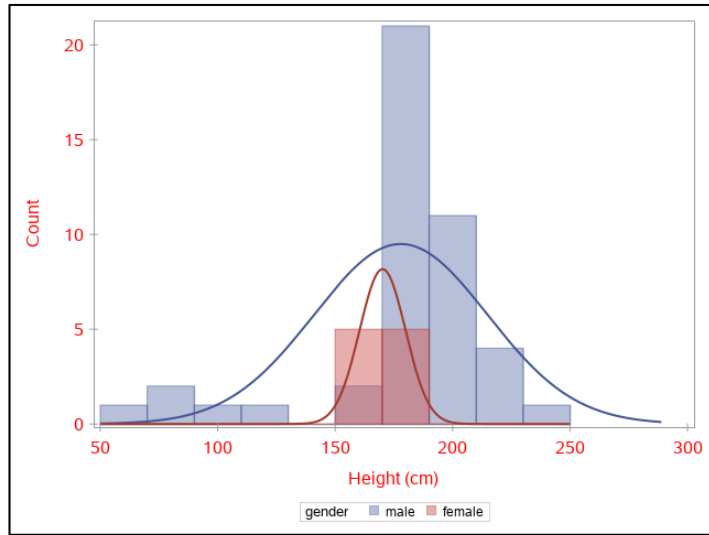
## I. Simple histogram upgraded

```
PROC SGPLOT data=Nile;
  histogram Nile/ fillattrs=(color="blue") nbins=6;
  xaxis values=(0,250,500,750,1000,1250,1500)
  label="Nile River flow rate";
```



## II. Two-way histogram upgraded

```
PROC SGPLOT data=starwars;
  histogram height/ group=gender transparency=0.5
  nbins=10 scale=count;
  density height/ group=gender type=normal;
  xaxis valueattrs=(size=12pt color="red") labelattrs=(size=12pt
  color="red") tickstyle=across label="Height (cm)";
  yaxis valueattrs=(size=12pt color="red") labelattrs=(size=12pt
  color="red") tickstyle=across;
```

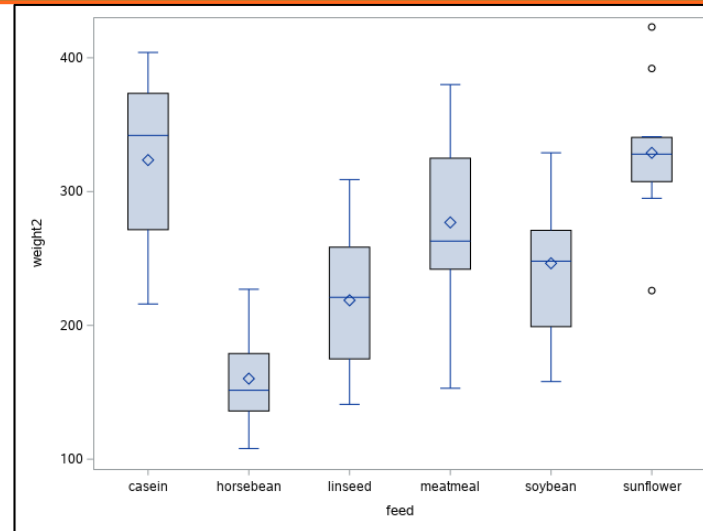


<https://create.kahoot.it/share/mmgg-in-sas-quick-test-1/cb125dae-c9f2-49dd-92b5-c0400a12ce2d>

# Boxplots

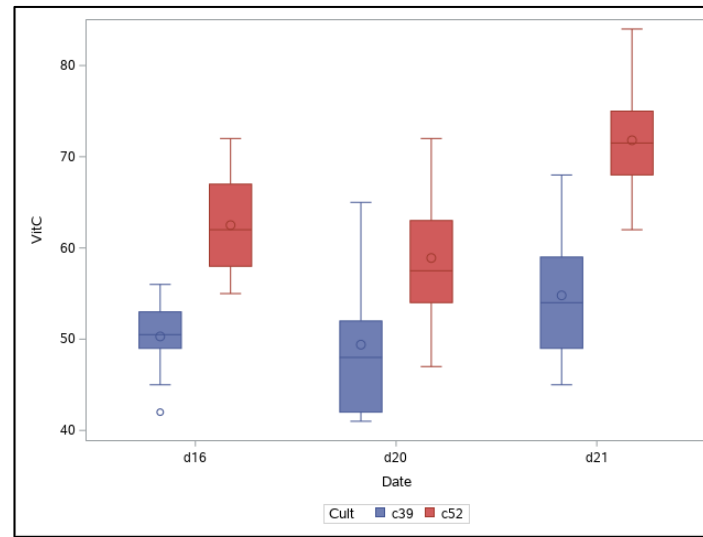
## I. Simple boxplot

```
PROC SGPLOT data=chickwts;  
  vbox weight2 /category=feed;
```



## II. Two-way boxplot

```
PROC SGPLOT data=cabbages;  
  vbox VitC /category=Date group=Cult;
```



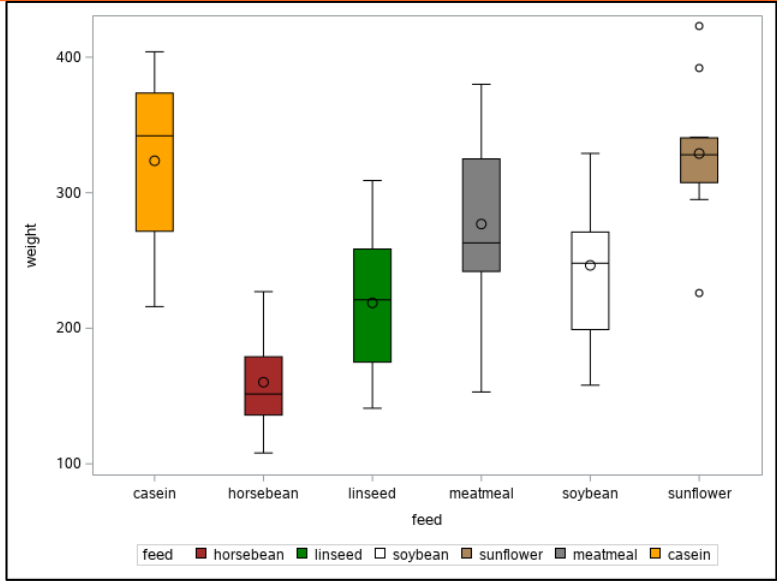
# Boxplots cont.

## I. Simple boxplot upgraded

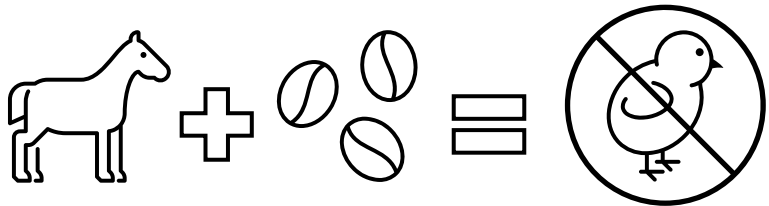
DATA chickwtsAttrMap;

length id \$5 value \$10 FillColor \$10;  
 input id value FillColor;  
 datalines;

chkID casein Orange  
 chkID horsebean Brown  
 chkID linseed Green  
 chkID meatmeal Grey  
 chkID soybean White  
 chkID sunflower Yellow  
 ;



```
PROC SGPLOT data=chickwts dattrmap=chickwtsAttrMap;
  vbox weight2 /category=feed group=feed boxwidth=0.50
  whiskerattrs=(color="black")
  lineattrs=(color="black")
  attrid=chkID
  medianattrs=(color="black")
  meanattrs=(color="black")
  outlierattrs=(color="black");
  yaxis label="weight";
```



# Boxplots cont. 2

## II. Two-way boxplot upgraded

DATA cabbagesAttrMap;

```
length id $5 value $3 FillColor $10 LineColor $10
      MarkerColor $10 MarkerSymbol $15
      MarkerSize 3;
```

```
input id value FillColor LineColor MarkerColor
      MarkerSymbol MarkerSize;
```

datalines;

```
cabID c39 Green Black BILG CircleFilled 6
cabID c52 Purple Black VLIP CircleFilled 6
;
```

DATA cabbages; set cabbages;

```
if Date="d16" then do Date2="Day 16"; end;
else if Date="d20" then do Date2="Day 20"; end;
else do Date2="Day 21"; end;
```

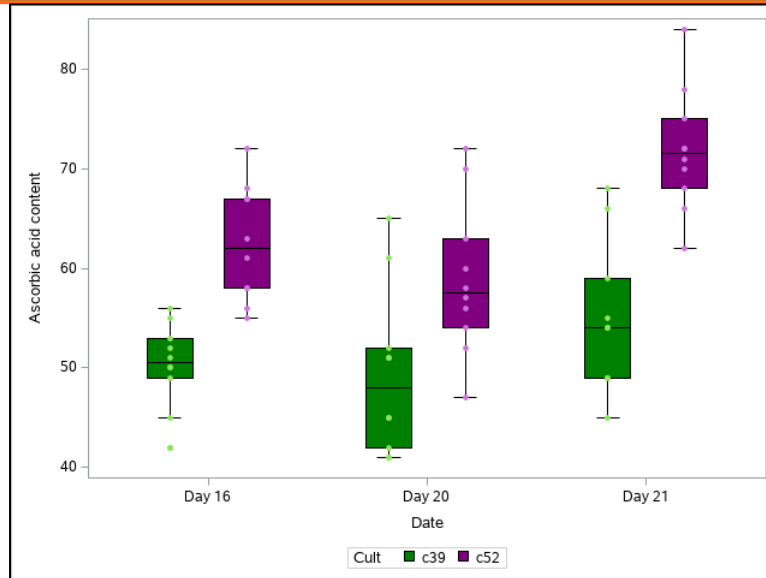
PROC SGPLOT data=cabbages dattrmap=cabbagesAttrMap;

```
vbox VitC /category=Date2 group=Cult attrid=cabID nomean;
```

```
scatter y=VitC x=Date2/ group=Cult groupdisplay=cluster clusterwidth=0.70 attrid=cabID;
```

```
yaxis label="Ascorbic acid content";
```

```
xaxis label="Date";
```



**Exploration:** try creating  
a two-way box plot of  
cabbage head weight  
(HeadWt) across Cultivar  
(Cult) and Date (Date)



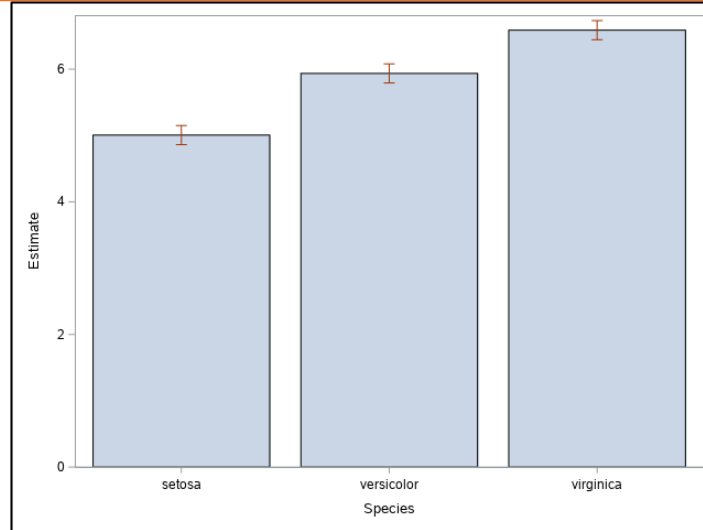


# Bar plots



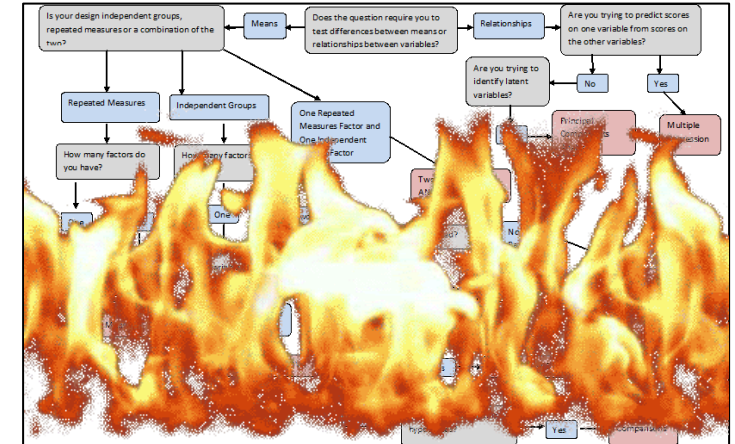
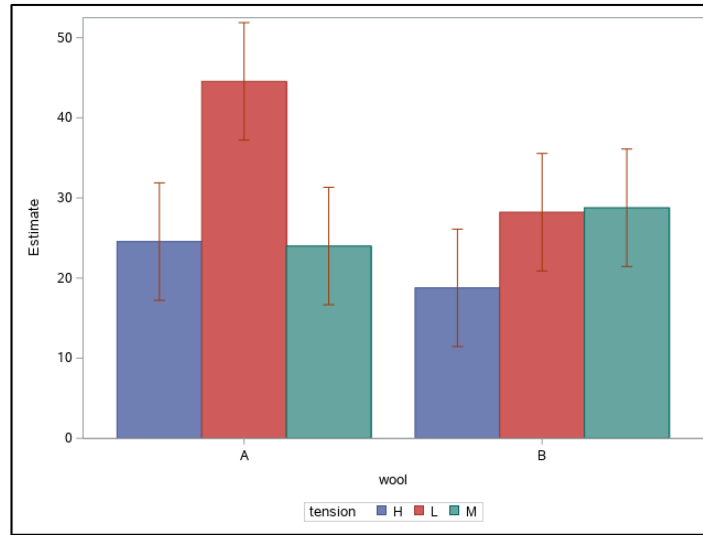
## I. Simple bar plot

```
PROC GLIMMIX data=iris;
  class Species;
  model SepalLength=Species;
  lsmeans Species / cl;
  ods output LSMeans=iris_means;
PROC SGPLOT data=iris_means;
  vbarparm category=Species response=Estimate /
  limitlower=Lower limitupper=Upper;
```



## II. Two-way bar plot

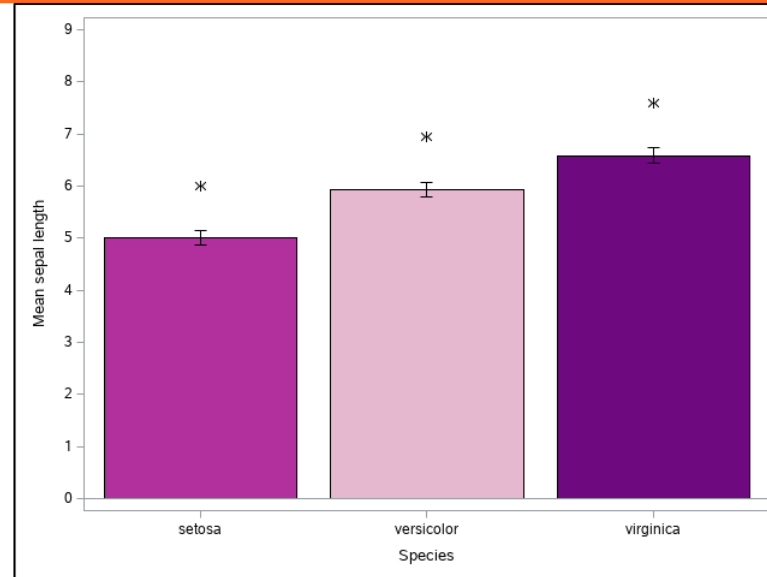
```
PROC GLIMMIX data=warpbreaks;
  class wool tension;
  model breaks=wool*tension;
  lsmeans wool*tension / cl;
  ods output LSMeans=wb_means;
PROC SGPLOT data=wb_means;
  vbarparm category=wool response=Estimate / group=tension
  groupdisplay=cluster
  limitlower=Lower limitupper=Upper;
```



# Bar plots cont.

## I. Simple bar plot upgraded

```
DATA irisAttrMap;
  length id $7 value $12 FillColor $10;
  input id value FillColor;
  datalines;
      irisID setosa STPPK
      irisID versicolor PAPPK
      irisID virginica VIP
  ;
DATA iris_means; set iris_means;
  Estimate2 = Estimate + 1;
```

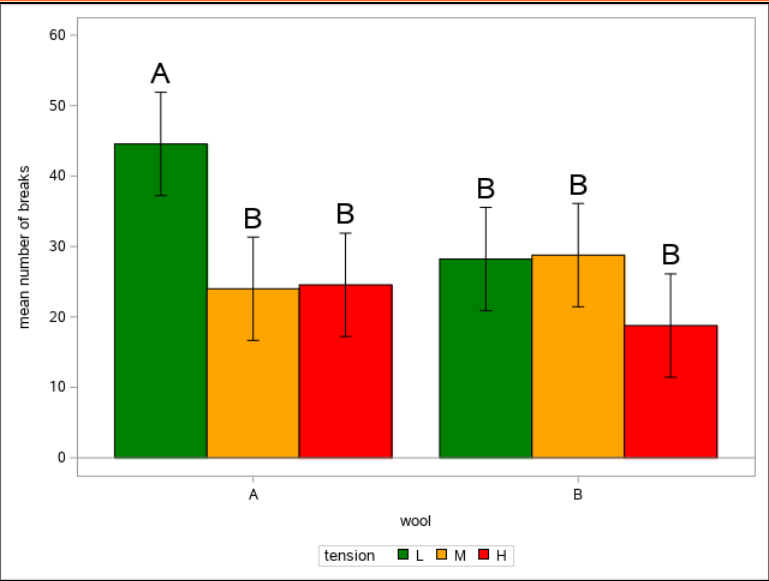


```
PROC SGPLOT data=iris_means dattrmap=irisAttrMap noautolegend;
  vbarparm category=Species response=Estimate/group=Species attrid=irisID;
  vbarparm category=Species response=Estimate/limitlower=Lower
    limitupper=Upper limitattrs=(color=black) nofill;
  yaxis label="Mean sepal length" values=(0,1,2,3,4,5,6,7,8,9);
  scatter y=Estimate2 x=Species/ markerattrs=(color=black size=10
    symbol=Asterisk);
```

# Bar plots cont. 2

## II. Two-way bar plot upgraded

```
DATA wbAttrMap;
  length id $4 value $1 FillColor $10 LineColor $10;
  input id value FillColor LineColor;
  datalines;
    wbID L Green Black
    wbID M Orange Black
    wbID H Red Black
  ;
DATA wb_means; set wb_means;
  if tension="L" then do tension2="1"; end;
  else if tension="M" then do tension2="2"; end;
  else do tension2="3"; end;
  if tension="L" and wool="A" then do text="A"; end;
  else do text="B"; end;
  Estimate2 =Estimate + 10;
PROC SORT data=wb_means; by wool tension2;
```



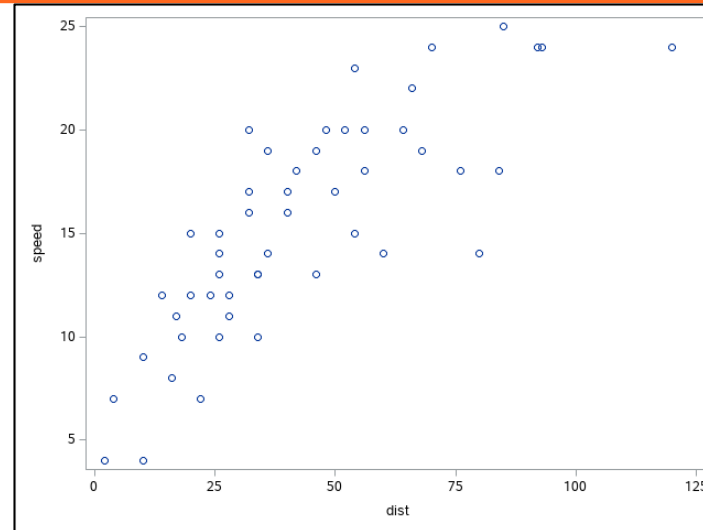
<https://app.animaker.com/animo/DtTiTnEg5y3oWQw/>

```
PROC SGPLOT data=wb_means dattrmap=wbAttrMap;
  vbarparm category=wool response=Estimate/group=tension groupdisplay=cluster
  limitlower=Lower limitupper=Upper
  attrid=wbID limitattrs=(color=black);
  yaxis label="mean number of breaks" values=(0,10,20,30,40,50,60);
  text Y=Estimate2 x=wool text=text/ group=tension groupdisplay=cluster
  textattrs=(size=18 color="black");
```

# Scatter plots

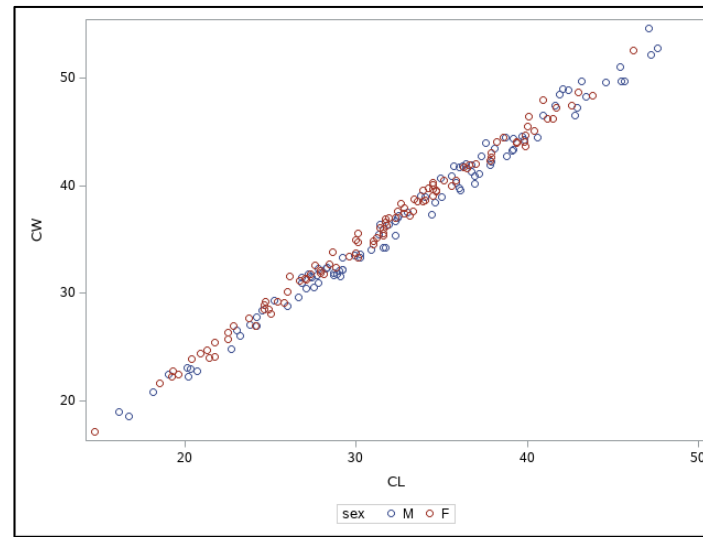
## I. Simple scatter plot

```
PROC SGPLOT data=cars;  
scatter y=speed x=dist;
```



## II. Two-sample scatter plot

```
PROC SGPLOT data=crabs;  
scatter y=CW x=CL / group=sex;
```



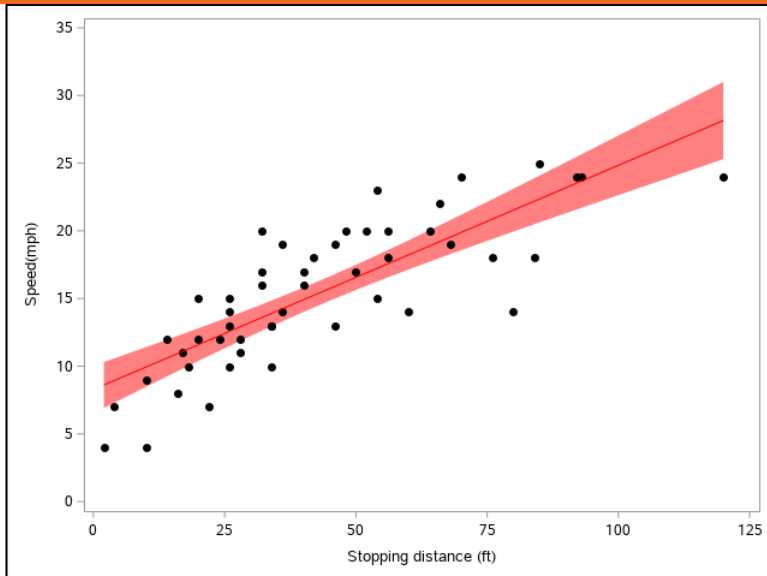
# Scatter plots cont. 2

## I. Simple scatter plot upgraded

```
PROC GLIMMIX data=cars;  
    model speed=dist;  
    output out=cars_pred pred lcl ucl;  
PROC SORT data=cars_pred; by dist;
```

**Pitfall  
Alerts**

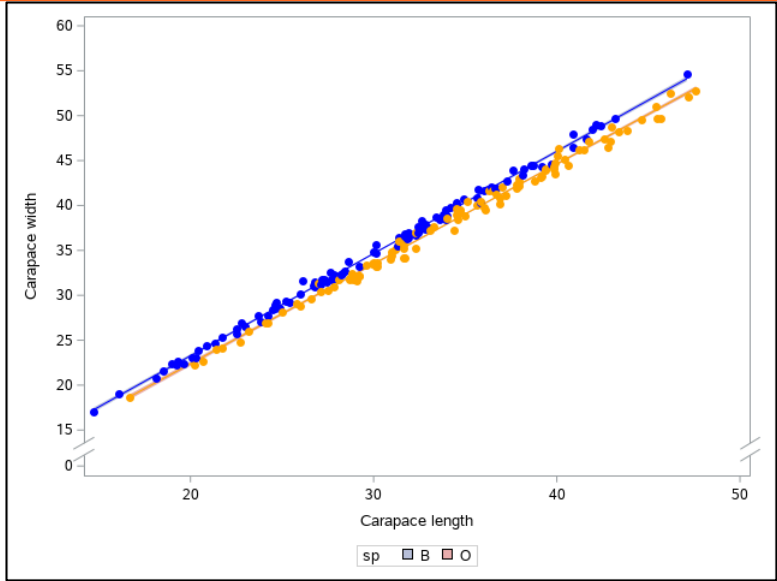
```
PROC SGPLOT data=cars_pred noautolegend;  
    band x=dist lower=lcl upper=ucl /transparency=0.50  
        fillattrs=(color='red');  
    scatter y=speed x=dist /  
        markerattrs=(symbol=circlefilled  
                    color="black");  
    series y=Pred x=dist / lineattrs=(color="red");  
    xaxis label=" Stopping distance (ft)";  
    yaxis label=" Speed (mph)";  
    values=(0,5,10,15,20,25,30,35);
```



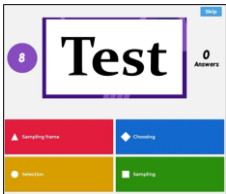
# Scatter plots cont.

## II. Two-sample scatter plot upgraded

```
PROC GLIMMIX data=crabs;
  class sp sex;
  model CW=CL|sp;
  output out=crabs_pred pred lcl ucl;
PROC SORT data=crabs_pred; by CL sp;
DATA crabAttrMap;
  length id $6 value $1 MarkerColor $10 LineColor $10
         Bandcolor $10;
  input id value MarkerColor LineColor BandColor;
  datalines;
      crabID B Blue Blue Blue
      crabID O Orange Orange Orange
  ;
```



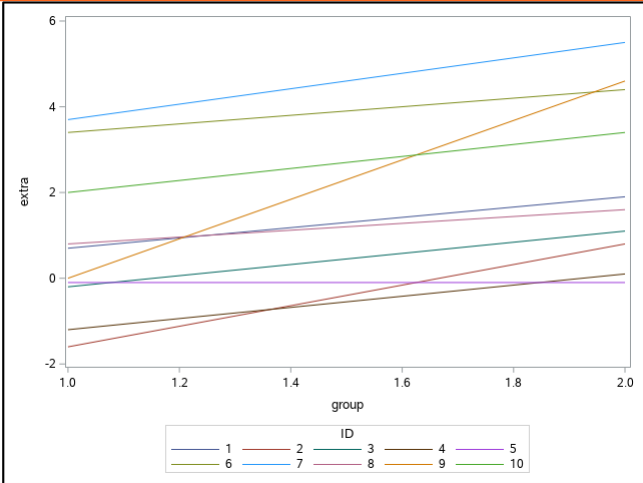
```
PROC SGPLOT data=crabs_pred dattrmap=crabAttrMap;
  band x=CL lower=lcl upper=ucl/group=sp transparency=0.50 attrid=crabID;
  series y=Pred x=CL/ group=sp attrid=crabID;
  scatter y=CW x=CL/ group=sp markerattrs=(symbol=circlefilled) attrid=crabID;
  xaxis label="Carapace length";
  yaxis label="Carapace width" ranges=(0-1 14-60) values=(0,15,20,25,30,35,40,45,50,55,60);
  styleattrs axisbreak=slantedright;
```



# Other plots

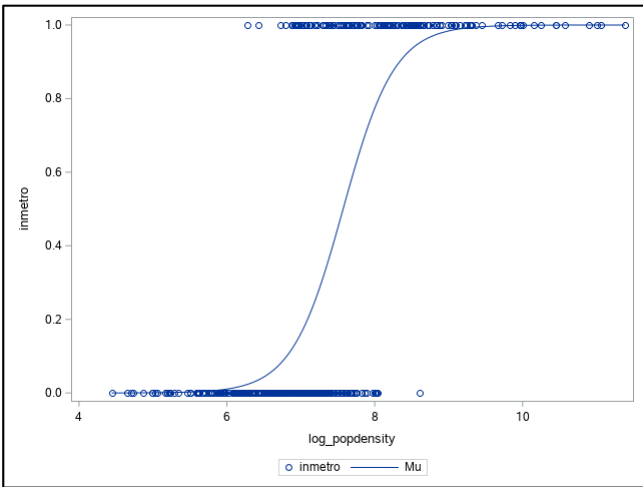
## I. Spaghetti plot

```
PROC SGPLOT data=sleep;
  series x=group y=extra / group=ID;
```



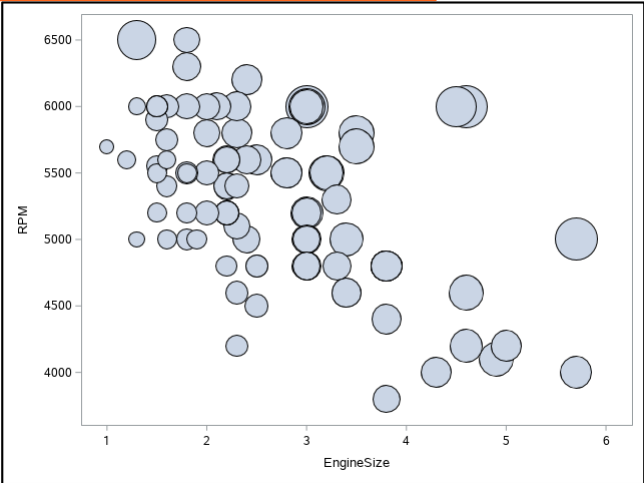
## II. Logistic regression plot

```
DATA midwest; set midwest;
  log_popdensity=log(popdensity);
PROC GLIMMIX data=midwest;
  model inmetro(event="1")=log_popdensity /dist=binary;
  output out=midwest_pred pred(ilink) lcl(ilink) ucl(ilink);
PROC SORT data=midwest_pred; by log_popdensity;
PROC SGPLOT data=midwest_pred;
  scatter y=inmetro x=log_popdensity;
  series y=PredMu x=log_popdensity;
```



## III. Bubble plot

```
PROC SGPLOT data=Cars93;
  bubble x=EngineSize y=RPM size=Horsepower;
```



# Other plots cont.

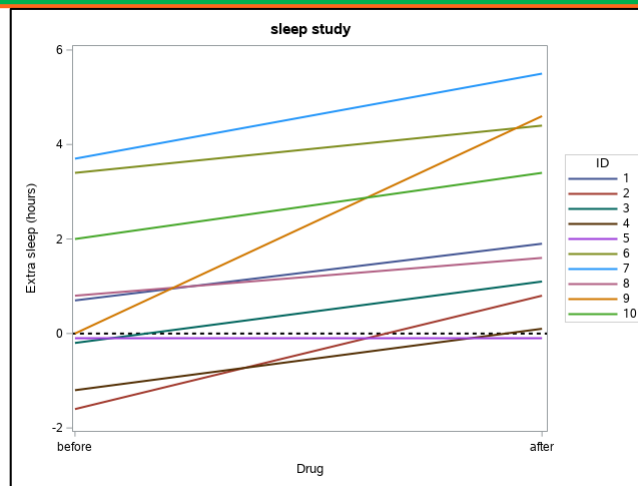
## I. Spaghetti plot upgrade

```
DATA sleep; set sleep;
```

```
if group=1 then do group2="before"; end;  
else do group2="after"; end;
```

```
PROC SGPLOT data=sleep;
```

```
title "sleep study";  
refline 0 / lineattrs=(thickness=2 color="black" pattern=ShortDash);  
series x=group2 y=extra / group=ID lineattrs=(thickness=2);  
yaxis label="Extra sleep (hours)";  
xaxis label="Drug";  
keylegend / location=outside position=right across=1;
```



## II. Logistic regression plot upgrade

```
PROC SGPLOT data=midwest_pred noautolegend;
```

```
band x=log_popdensity lower=LCLMu upper=UCLMu /  
transparency=0.5 fillattrs=(color="grey");
```

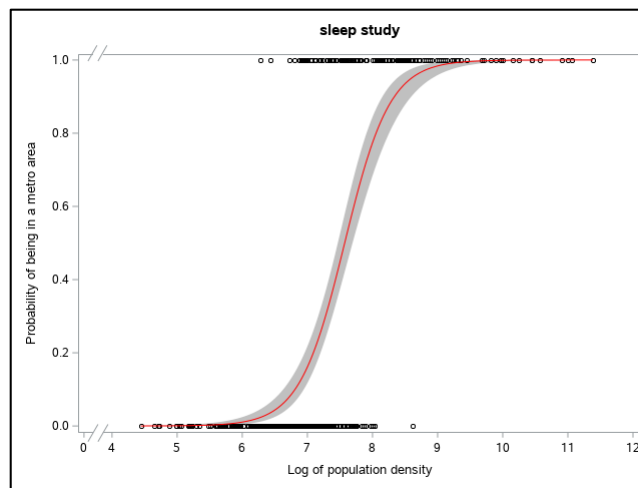
```
scatter y=inmetro x=log_popdensity /  
markerattrs=(size=6 color="black");
```

```
series y=PredMu x=log_popdensity / lineattrs=(color="red");
```

```
yaxis label="Probability of being in a metro area";
```

```
xaxis label="Log of population density" ranges=(0-0.1 3.9-12)  
values=(0,4,5,6,7,8,9,10,11,12);
```

```
styleattrs axisbreak=slantedright;
```





# Other plots cont. 2

## III. Bubble plot upgrade

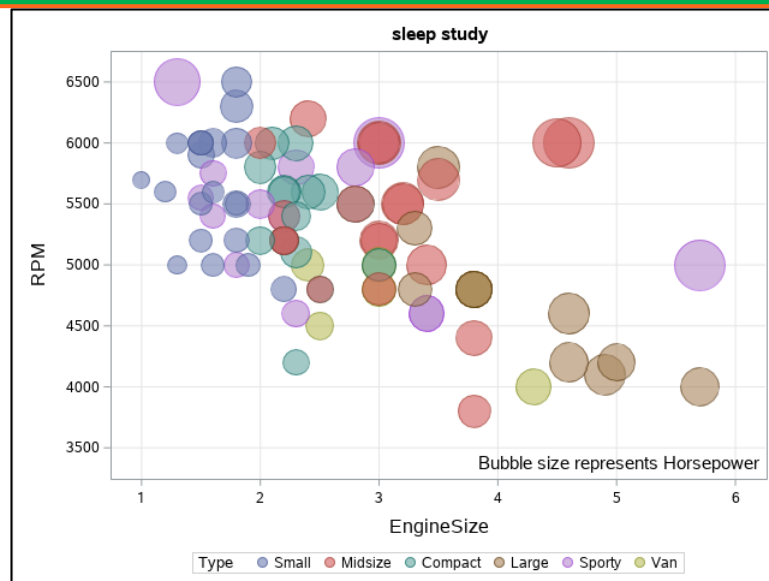
```
PROC SGPLOT data=Cars93;
```

```
  bubble x=EngineSize y=RPM size=Horsepower/  
        group=Type transparency=0.4;
```

```
  inset "Bubble size represents Horsepower" /  
        position=bottomright textattrs=(size=11);
```

```
  yaxis grid values=(3500,4000,4500,5000,5500,6000,6500)  
        labelattrs=(size=12);
```

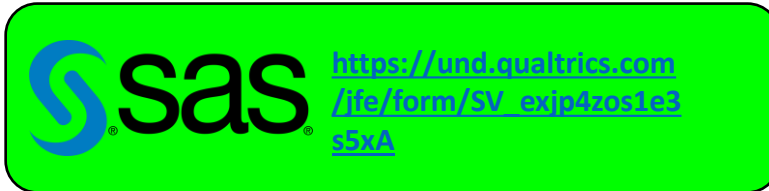
```
  xaxis grid labelattrs=(size=12);
```



**Exploration:** try creating a couple different attribute maps and try it out on the sleep study bubble plot

# Closing

- Please try out the post-test and survey

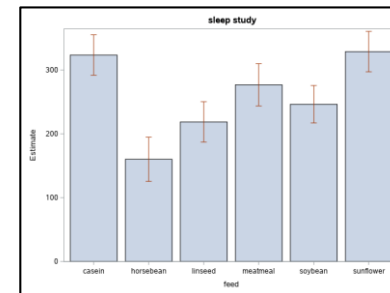
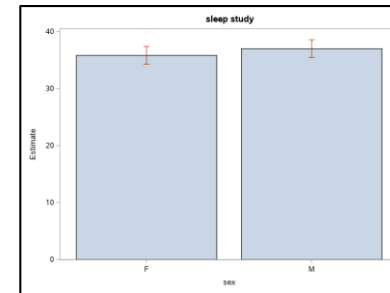


- **Special Treat:** Example SAS-code contains a basic introduction to macros that allow you to generate statistics and prebuilt graphs!

```
%mean_test
```

```
%plot_tests
```

- You can find the macro code and examples at the bottom of the SAS-code



<https://app.animaker.com/animoyLEBSrgOcYo6aX9g/>

# References

- ✓ [https://www.lexjansen.com/wuss/2015/141\\_Final\\_Paper\\_PDF.pdf](https://www.lexjansen.com/wuss/2015/141_Final_Paper_PDF.pdf)
- ✓ <https://blogs.sas.com/content/iml/2018/12/03/tips-customize-legends-proc-sgplot.html>
- ✓ <https://blogs.sas.com/content/iml/2012/10/17/specify-the-colors-of-groups-in-sas-statistical-graphics.html>
- ✓ <https://communities.sas.com/t5/Graphics-Programming/SGPLOT-VBOX-Change-Category-Color/td-p/425262>
- ✓ <https://communities.sas.com/t5/Graphics-Programming/SGPLOT-VBOX-Change-Category-Color/td-p/608403>
- ✓ [https://documentation.sas.com/?cdclid=pgmsascdc&cdcVersion=9.4\\_3.5&docsetId=grstatproc&docsetTarget=n18szqcwir8q2nn10od9hhdh2ksj.htm&locale=en](https://documentation.sas.com/?cdclid=pgmsascdc&cdcVersion=9.4_3.5&docsetId=grstatproc&docsetTarget=n18szqcwir8q2nn10od9hhdh2ksj.htm&locale=en)
- ✓ [https://support.sas.com/content/dam/SAS/support/en/books/pro-template-made-easy-a-guide-for-sas-users/62007\\_Appendix.pdf](https://support.sas.com/content/dam/SAS/support/en/books/pro-template-made-easy-a-guide-for-sas-users/62007_Appendix.pdf)
- ✓ <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3644-2019.pdf>
- ✓ <https://support.sas.com/kb/52/964.html>
- ✓ <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2179-2018.pdf>
- ✓ <https://blogs.sas.com/content/graphicallyspeaking/2017/12/19/getting-started-sgplot-part-9-bubble-plot/>
- ✓ [https://documentation.sas.com/?cdclid=pgmsascdc&cdcVersion=9.4\\_3.5&docsetId=grstatproc&docsetTarget=p0er4dg9tojp05n1sf7maeqdz1d8.htm&locale=en](https://documentation.sas.com/?cdclid=pgmsascdc&cdcVersion=9.4_3.5&docsetId=grstatproc&docsetTarget=p0er4dg9tojp05n1sf7maeqdz1d8.htm&locale=en)
- ✓ <https://blogs.sas.com/content/iml/2018/12/03/tips-customize-legends-proc-sgplot.html>
- ✓ <https://support.sas.com/rnd/datavisualization/yourGraphs/businessQuick/bubble/>

# Acknowledgements



- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

**DaCCoTA**  
DAKOTA CANCER COLLABORATIVE  
ON TRANSLATIONAL ACTIVITY