

## A Gentle Introduction to Creating SAS Graphs, Part 2

Dane Korver, RTI Health Solutions, RTP, NC

### ABSTRACT

This paper is a continuation of my previous SAS paper, "A Gentle Introduction to Creating SAS Graphs." In my previous paper, we discussed the five ODS Graphics procedures, how to create a simple bar chart and outputting your graphs. In this paper, we are going to continue using my Wordle data to keep practicing our SAS graphing skills.

### INTRODUCTION

The goal of creating a graph is to help us visualize our data that is intuitive, easy to read and provides that extra insight that we cannot get by reading a table. In this paper, we are going to continue creating SAS graphs that go beyond the basics using my Wordle data. I have been keeping a record of my daily plays in a spreadsheet that contains two variables: **Date** and **Number\_of\_Guesses**. For people who are not familiar with Wordle, you have six tries to guess the daily word. For days where I have not been able to guess the daily Wordle, I denoted those with a "7." We are going to focus our attention on enhancing a similar side-by-side bar chart that I created last year and have a little fun creating a "waterfall" type plot to see if we might be able to find a "winning streak" between Player's.

### A SIDE-BY-SIDE BAR CHART WITH ENHANCEMENTS

In my previous paper, I ended with an example of a side-by-side bar chart that compared the distribution of Wordle guesses between players. That was a nice start, but we can do better! A key piece of information that probably should have been added to the legend is how many days did each Player play? If one player didn't play for a week and the other player did, that might show a shift in the distribution. To add this to the legend, I used **PROC FREQ** to find the Player counts for the month, created a variable named **Player2** that combined the player's name and their counts so it could be in a nice format, merged it

back to my main **Wordle** dataset and created macros to use in my attribute map for the graph legend.

```
*Obtain Player counts;
proc freq data=wordle NOPRINT;
where MISSING(Number_of_Guesses)=0;
table Player/out=summary;
run;

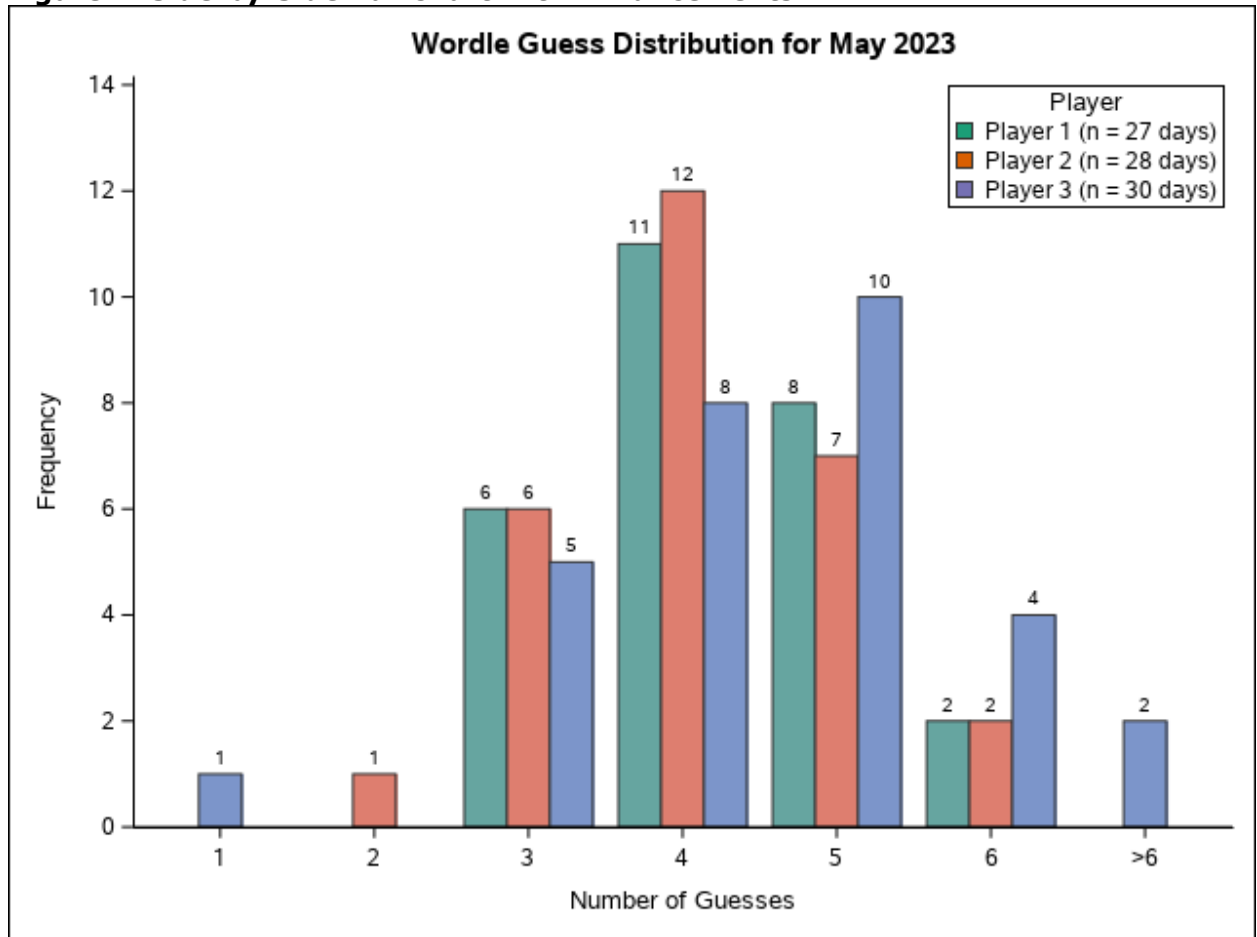
*Formatting;
data summary(keep=Player COUNT Player2);
set summary;
length Player2 $50.;
Player2 = STRIP(Player)||' (n = '||compress(put(COUNT,8.))||)';
run;

*Merge COUNTS to Wordle data;
data wordle;
merge wordle summary;
by Player;
run;

*Create macros;
data _NULL_;
set summary;
IF Player="Player 1" then call symput("player1",STRIP(COUNT));
IF Player="Player 2" then call symput("player2",STRIP(COUNT));
IF Player="Player 3" then call symput("player3",STRIP(COUNT));
run;

*Create Graph;
data myattrs;
length value $50. fillcolor $20;
show='AttrMap';
id="some_id";
value="Player 1 (n = &player1. days)"; fillcolor="#1b9e77"; output;
value="Player 2 (n = &player2. days)"; fillcolor="#d95f02"; output;
value="Player 3 (n = &player3. days)"; fillcolor="#7570b3"; output;
run;
ODS GRAPHICS ON /reset = all imagename="Side-by-side plot" imagefmt=png;
ODS LISTING GPATH = "<file address to putput your PNG file>";
title 'Wordle Guess Distribution for May 2023';
proc sgplot data=wordle dattrmap=myattrs noborder;
vbar Number_of_Guesses/group=Player2 groupdisplay=cluster attrid=some_id
datalabel outlineattrs=(color=gray33);
xaxis values=(1 2 3 4 5 6 7) valuesdisplay=("1" "2" "3" "4" "5" "6" ">6")
label="Number of Guesses";
yaxis values=(0 to 14 by 2);
keylegend/title="Player" across=1 position=topright location=inside;
run;
quit;
ODS GRAPHICS OFF;
```

**Figure 1. Side-by-Side Bar Chart with Enhancements**



This graph now gives a nice look at how many days during the month each Player played and their guess distribution. To feed that “competitive” spirit, let us see if we can create a figure that graphs the “win margin” for each day during the month and see if there are any particular “win streaks” for a particular player. To do this, I created a “waterfall type” plot, but instead of plotting the difference between the “highest number of guesses” and the player with the “lowest number of guesses”, I wanted to graph the “spread” between the “highest number of guesses” and “lowest number of guesses” and then color code by Player.

## WATERFALL TYPE PLOT

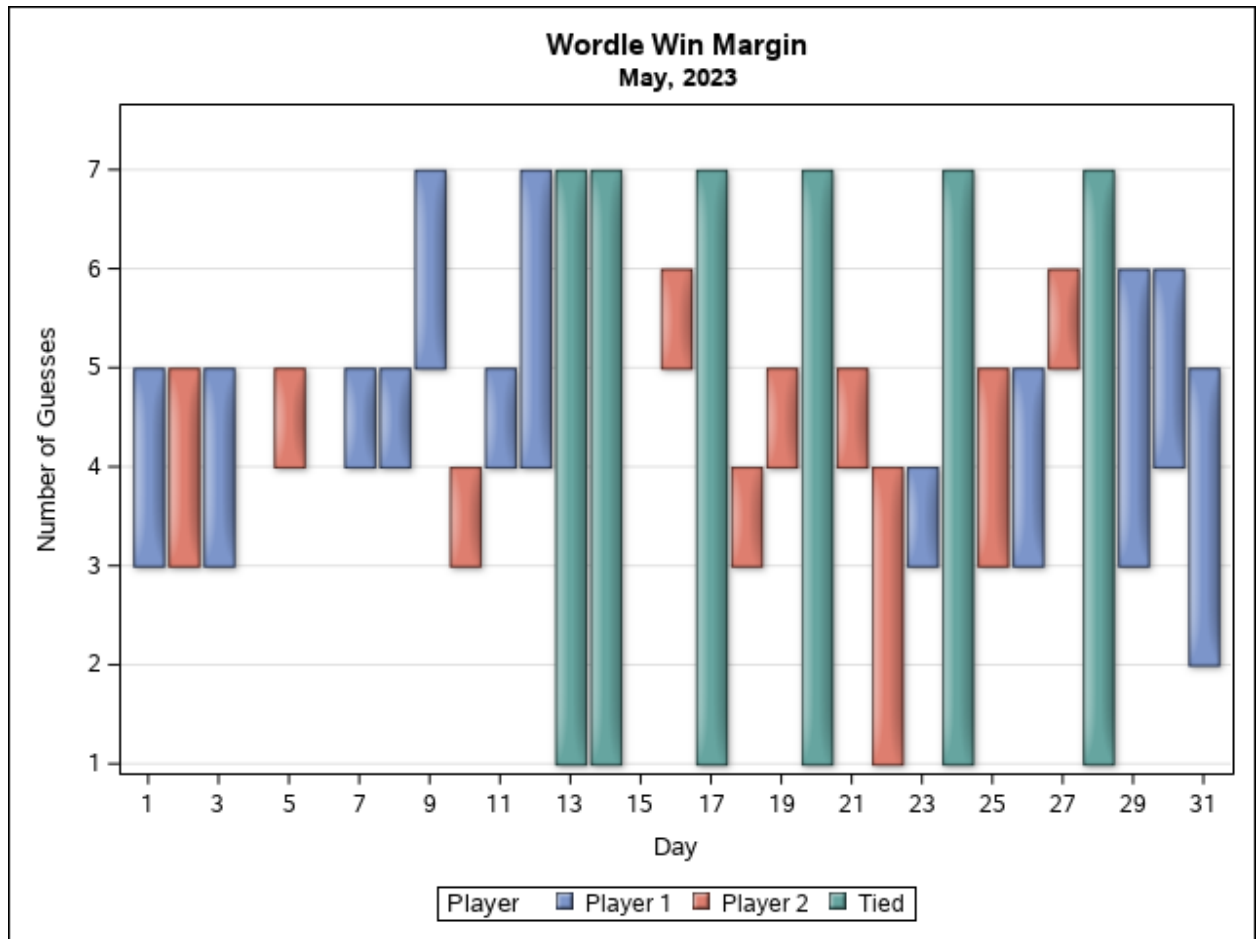
To create a “waterfall” plot, we will continue to use PROC SGPLOT, but instead of using the **VBAR** statement, we will use the **HIGHLOW** statement. To get the data in the format we need, I did some data management magic and created a dataset with the following variables: **Date**, **Player1\_Score**, **Player2\_Score**, and a variable named **Player** that indicated who had the lowest number of guesses and computed the absolute value of the **Difference** between the **Player1\_Score** and **Player2\_Score** variables. I did this in Excel since that is where I had my data and it was a small dataset. After reading my data into SAS, I created a variable named **winning\_score** for the Player who had the fewest number of guesses and **losing\_score** for the Player who had the most number of guesses. For days where we “Tied,” I set **winning\_score**=1 and **losing\_score**=7 because I wanted to see a “full bar” in the graph so we could easily see the days we “Tied.” The next page shows the code used to create this graph.

```

data wordle_margin;
format Player $50.;
set wordle_margin;
Difference=Difference*-1;
day+1;
if NMISS(Player1_Score,Player2_Score)=0 then do;
    if Player1_Score<Player2_Score then do;
        winning_score=Player1_Score;
        losing_score=Player2_Score;
    end;
    if Player1_Score>Player2_Score then do;
        winning_score=Player2_Score;
        losing_score=Player1_Score;
    end;
end;
if Player="Tied" then do;
    winning_score=1;
    losing_score=7;
end;
run;
ODS GRAPHICS ON /reset = all imagename="plot3" imagefmt=png;
ODS LISTING GPATH = "/sasdata/hsdevdata1/OPQtesting/dkorver/SESUG/YR2023/";
title1 "Wordle Win Margin";
title2 "May, 2023";
proc sgplot data=wordle_margin;
highlow x=Day low=winning_score high=losing_score / type=bar group=Player
    lineattrs=graphoutlines dataskin=matte fillattrs=(transparency=0);
xaxis label="Day" values=(1 to 31 by 1);
yaxis label="Number of Guesses" grid offsetmax=0.1;
run;
quit;
ODS GRAPHICS OFF;

```

**Figure 2. Waterfall Type Plot**



This graph does a good job of showing which Player had a "Win Streak," and shows the margin between the "winner" and "loser" for each day in May.

## CONCLUSION

I hope this paper illustrated a couple of examples where doing a little data management magic can help get what you are desiring for your graph even if there isn't an easy way to get it done directly thru PROC SGPLOT and its style options. Sometimes a little creativity goes a long way with creating your desired graph!

## ACKNOWLEDGMENTS

This paper is dedicated to the people that I play Wordle with every morning.

## RECOMMENDED READING

- Checkout SAS's "Graphically Speaking" blog here: <https://blogs.sas.com/content/graphicallyspeaking/> or "The DO Loop" blog here: <https://blogs.sas.com/content/iml/>
- Harris, Kriss, and Richann Watson. 2020. *SAS Graphics for Clinical Trials by Example*. Cary, NC: SAS Institute Inc.

## **CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Dane Korver  
RTI Health Solutions  
dkorver@rti.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.