**proc** **import** file = "C:\Users\dryan\OneDrive\Documents\Sts499\president\_polls\_2020\_generalElection.csv" out = polls2020

dbms = csv replace;

datarow = **2**;

getNames = Yes;

**run**;

**data** polls2020;

set polls2020;

if answer = 'Biden' then porp\_biden = pct;

if answer = 'Trump' then porp\_trump = pct;

**run**;

**proc** **freq** data = polls2020;

tables answer;

**run**;

**data** biden;

set polls2020;

where answer in('Biden');

\*if cmiss(state) > 0 then delete;

**run**;

**data** biden;

set biden;

diff\_time = election\_date - end\_date;

**run**;

**proc** **means** data = biden;

class state;

var porp\_biden;

**run**;

**data** biden;

set biden;

rename fte\_grade = grade;

**run**;

**proc** **sort** data = biden;

by grade;

**run**;

**data** biden;

set biden;

if cmiss(grade) > **0** then delete;

**run**;

**data** biden;

set biden;

if (grade = 'A+' | grade = 'A' | grade = 'A-' | grade = 'A/' | grade = 'B+') then new\_grade = 'A/B';

if(grade = 'B' | grade = 'B-' | grade = 'B/' | grade = 'C+' | grade = 'C' | grade = 'C-' | grade = 'C/' | grade = 'D-') then new\_grade = 'B/C';

**run**;

**proc** **sort** data = biden;

by state;

**run**;

\*ods pdf file = 'C:\Users\dryan\OneDrive\Documents\Sts499\statePlots20\_biden.pdf';

**proc** **sgplot** data = biden ;

where new\_grade in("A/B","B/C");

scatter x = diff\_time y = porp\_biden/group = new\_grade;

\*band x = diff\_time upper = upper\_two\_week lower = lower\_two\_week / transparency = 0.5;

\*lineparm x= state\_p y = state\_p slope = 0;

xaxis reverse;

title 'all states loess 2020';

loess x = diff\_time y = porp\_biden/ nomarkers group = new\_grade;

by state;

**run**;

\*ods pdf close;

**proc** **sort** data = biden;

by state;

**run**;

**proc** **mixed** data = biden;

class new\_grade pollster\_id;

model porp\_biden = new\_grade / solution;

random int / subject = pollster\_id vcorr;

by state;

**run**;

**proc** **sort** data = biden;

by state;

**run**;

\*ods pdf file = 'C:\Users\dryan\OneDrive\Documents\Sts499\spaghetti20\_biden.pdf';

**proc** **sgplot** data = biden ;

where new\_grade in("A/B","B/C") /\*& diff\_time <101\*/;

series x = diff\_time y = porp\_biden/group = pollster groupLC = new\_grade markers groupLP = new\_grade groupMC = pollster groupMS = new\_grade;

\*band x = diff\_time upper = upper\_two\_week lower = lower\_two\_week / transparency = 0.5;

\*lineparm x= state\_p y = state\_p slope = 0;

xaxis reverse;

title 'spaghetti plot 2020';

\*loess x = diff\_time y = porp\_trump/ nomarkers group = new\_grade;

by state;

**run**;

\*ods pdf close;

**data** trump;

set polls2020;

where answer in('Trump');

**run**;

**data** trump;

set trump;

diff\_time = election\_date - end\_date;

**run**;

**data** trump;

set trump;

rename fte\_grade = grade;

**run**;

**data** trump;

set trump;

if (grade = 'A+' | grade = 'A' | grade = 'A-' | grade = 'A/' | grade = 'B+') then new\_grade = 'A/B';

if(grade = 'B' | grade = 'B-' | grade = 'B/' | grade = 'C+' | grade = 'C' | grade = 'C-' | grade = 'C/' | grade = 'D-') then new\_grade = 'B/C';

**run**;

**data** trump1;

set trump;

**run**;

**data** trump1;

set trump1;

where new\_grade in('A/B') | new\_grade in('B/C');

**run**;

**proc** **sort** data = trump1;

by state;

**run**;

**data** trump1;

set trump1;

diff\_time = election\_date - end\_date;

**run**;

\*ods pdf file = 'C:\Users\dryan\OneDrive\Documents\Sts499\statePlots20\_trump.pdf';

**proc** **sgplot** data = trump1 ;

scatter x = diff\_time y = porp\_trump/group = new\_grade;

\*lineparm x= state\_p y = state\_p slope = 0;

xaxis reverse;

title 'all states loess 2020';

loess x = diff\_time y = porp\_trump/ nomarkers group = new\_grade;

by state;

**run**;

\*ods pdf close;

**proc** **means** data = trump1;

class state;

var porp\_trump;

**run**;

**proc** **sort** data = trump1;

by state pollster\_id diff\_time;

**run**;

**data** polls2;

length fte\_grade $**4**;

set polls2020;

daysbefore = election\_date - end\_date;

if daysbefore <= **100** and state ne " " and population="lv"; \*keep only those within 100 days, in a state, with likely voters;

**run**;

**proc** **sort** data=polls2;

by state poll\_id question\_id; \*pollster daysbefore;

**run**;

**data** both;

merge polls2(where=(answer="Biden") rename=(pct=Bidenpct))

polls2(where=(answer="Trump") rename=(pct=Trumppct)); \*merge the Biden and Trump results onto one record, but note must rename the variable PCT;

by state poll\_id question\_id; \*pollster daysbefore;

if fte\_grade in ("A+", "A", "A-", "A/", "B+") then newgrade="A/B";

else if fte\_grade in ("B/", "B", "B-", "C+") then newgrade="B/C";

else if fte\_grade in ("C/", "C", "C-", "D-") then newgrade="C/D"; \* New grade group, to include SurveyMonkey mostly;

pBiden = Bidenpct/(Bidenpct + Trumppct);

pTrump = Trumppct / (Bidenpct + Trumppct);

keep state pollster daysbefore Bidenpct Trumppct fte\_grade population--office\_type end\_date election\_date

newgrade pbiden ptrump question\_id;

**run**;