

DM & Visualization - Exam 2020-01-31 - Solution

Version 1.1 - 2020-20-11

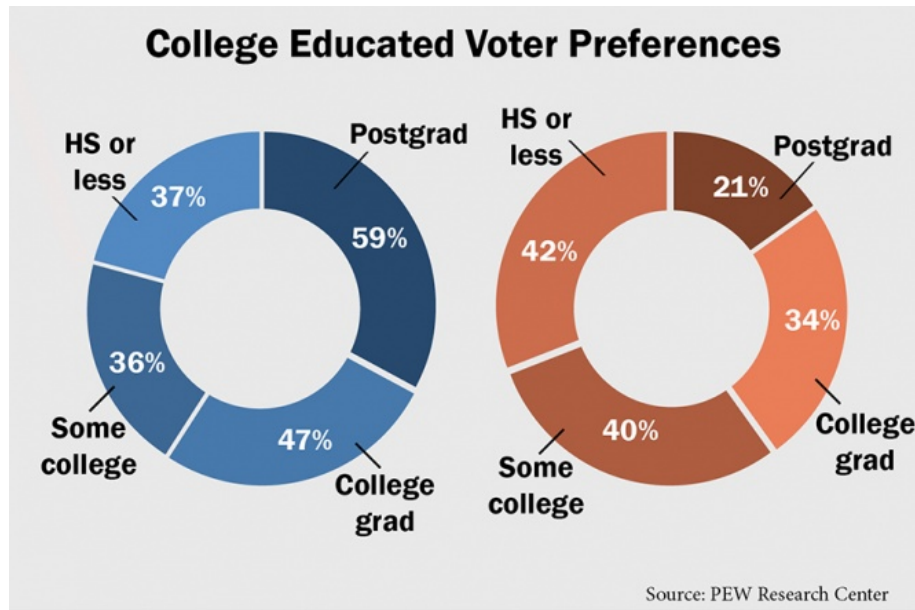


Figure 1: Source: <https://thedailytexan.com/2016/10/21/college-educated-voters-will-back-clinton-according-to-survey>

Analysis

Analyze the above graph that was published by "The Daily Texan" on Oct 21, 2016 in an article entitled "College-educated voters will back Clinton, according to survey".

Please remember that the conventional color for the Democratic Party in US is blue and for the Republican Party it is red.

Question: Is there one (or more) question addressed by the visualization?

The question is fairly clear: how does preference for the two party varies as educational level changes

Data: Is the data quality appropriate?

The quality of the data is reasonably suitable to answer the question.

- the percentages of each donut do not sum to 100% because they refer to different wholes,
- the values should be summed by education level, though they do not reach 100% but are around 80%, we can assume there is roughly 20% of "undecided".

Using the ISO 250112 Characteristics as guidelines we can identify two non adequate aspects:

Characteristic	Adequate	Comments
Accuracy	Yes	Percentage numbers
Completeness	Partly	Not sure about how to interpret the missing 20%
Consistency	Partly	Sum of percentages not equal to 100%
Currency	Yes	Presumably data is from 2016, but that is ok
Credibility	Yes	Source is reported as a polling company
Understandability	Yes	Data is easy to understand
Precision	Yes	Single % point precision is reasonable

Visual Proportionality: Are the values encoded in a uniformly proportional way?

The proportionality is completely altered by the wrong use of data: pie/donut MUST be used to represent part-whole

relationships only. In particular:

- % close to 50% are not half donut as one would expect
- sectors representing 59% and 42% across donuts have almost the same size
- moreover, areas and angles are generally not perceived accurately.

The educational levels are increasing (thus ordered) one would expect them to be encoded as color with increasing intensity, which is not the case.

Visual Utility: All the elements in the graph convey useful information?

All elements in the graph convey useful information. One might argue the thick lines separating the sectors could be removed, though colors are very similar and removing the lines could introduce a clarity issue. The background color is not really useful, but being uniform does not represent an issue.

Visual Clarity: Are the data in the graph clearly identifiable and understandable (properly described)?

The color coding relies on the implicit (for US citizens) associations of red to Republican party and blue to Democratic party. The labels are placed close and connected to the items they describe.

Design

Design the visualization based on the following data structure

Field	Dim./Measure	Description
EDUCATION_LEVEL	Dimension	Level of education of respondent
PARTY_PREFERENCE	Dimension	Party preferred by respondent, either Dem or Rep
PREFERENCES	Measure	Percentage of respondent expressing preference for that party

A full set of solutions (developed in Tableau) are available here:

<https://public.tableau.com/profile/mtorchiano/#!/vizhome/DMV2020-01-31VizExamSolution/DMV2020-01-31>

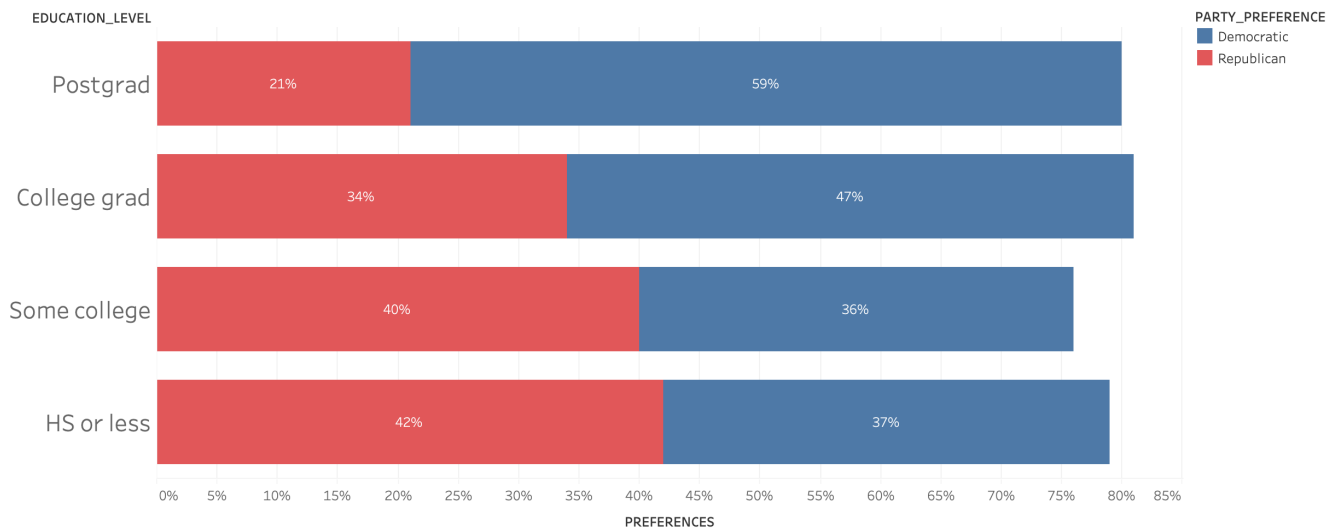
Design schema

A minimal solution reports the data as it is.

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	SUM(PREFERENCES)
Graph type:	Bar
Color:	PARTY_PREFERENCE
Size:	
Label:	

Sketch of the resulting graph:

Bar Chart Preferences Stacked

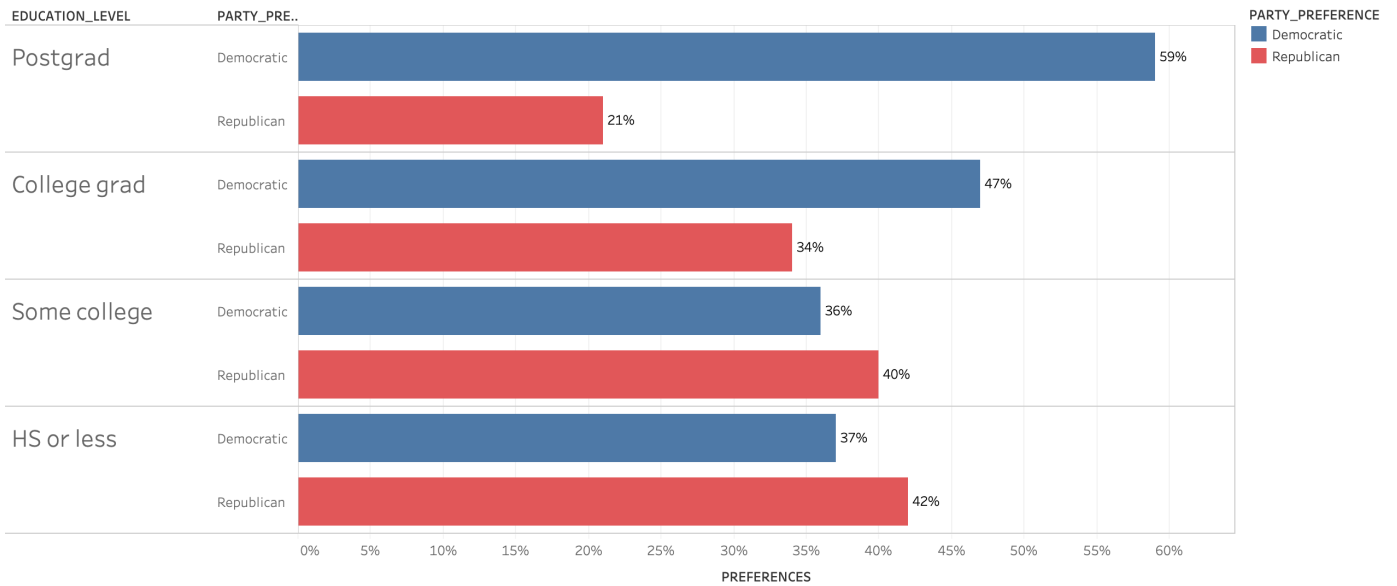


Sum of PREFERENCES for each EDUCATION_LEVEL . Color shows details about PARTY_PREFERENCE. The marks are labeled by sum of PREFERENCES.

Using bars not stacked but side by side makes it easier to perform meaningful comparison within each educational level.

Schema	Details
Rows:	EDUCATION_LEVEL, PARTY_PREFERENCE
Columns:	SUM(PREFERENCES)
Graph type:	Bar
Color:	PARTY_PREFERENCE
Size:	
Label:	PARTY_PREFERENCE

Bar Chart Preferences Sided

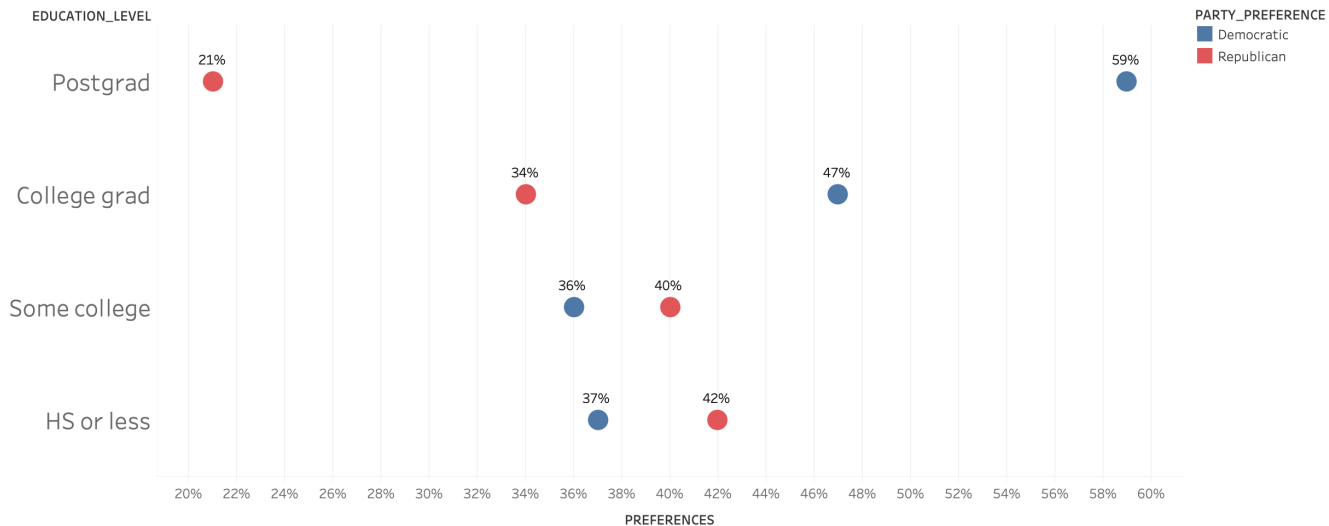


Sum of PREFERENCES for each PARTY_PREFERENCE broken down by EDUCATION_LEVEL . Color shows details about PARTY_PREFERENCE. The marks are labeled by sum of PREFERENCES.

Using dots instead of bars (no need to have a zero based scale on axis)

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	SUM(PREFERENCES)
Graph type:	Circle
Color:	PARTY_PREFERENCE
Size:	
Label:	PARTY_PREFERENCE

Dot Plot Preferences

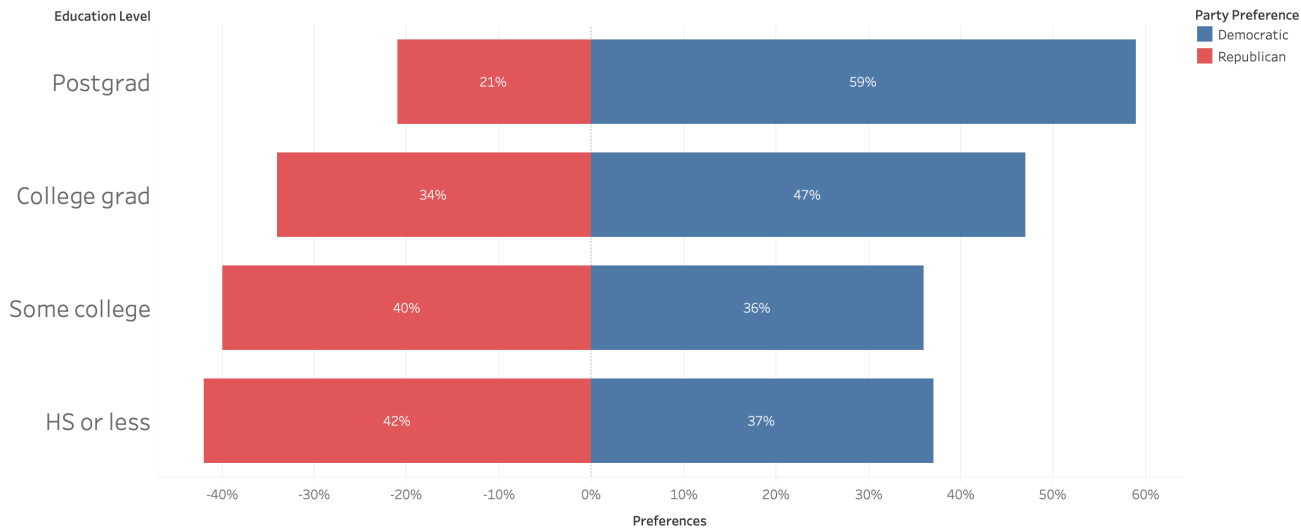


Sum of PREFERENCES for each EDUCATION_LEVEL . Color shows details about PARTY_PREFERENCE. The marks are labeled by sum of PREFERENCES.

Bars can be diverging.

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	SUM(iif(PARTY_PREFERENCE=="Democratic",1,-1)*[PREFERENCES])
Graph type:	Bar
Color:	PARTY_PREFERENCE
Size:	
Label:	PARTY_PREFERENCE

Bar Chart Preferences Stacked Diverging

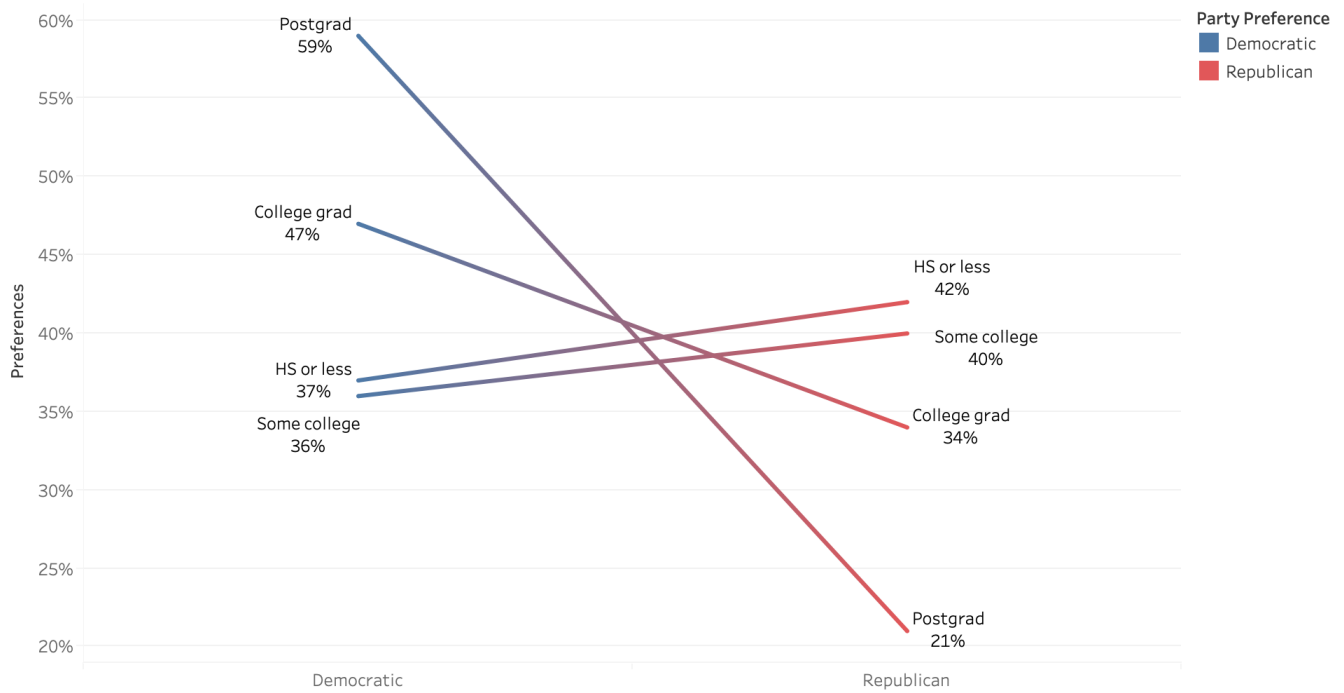


SUM(iif([Party Preference]=="Democratic",1,-1)*[Preferences]) for each Education Level. Color shows details about Party Preference. The marks are labeled by sum of Preferences.

Instead of bars we can use a slope chart to emphasize the change in orientation for different educational levels

Schema	Details
Rows:	SUM(Preferences)
Columns:	PARTY_PREFERENCE
Graph type:	Line
Color:	PARTY_PREFERENCE
Size:	
Label:	EDUCATION_LEVEL, PARTY_PREFERENCE

Slope Chart Preferences

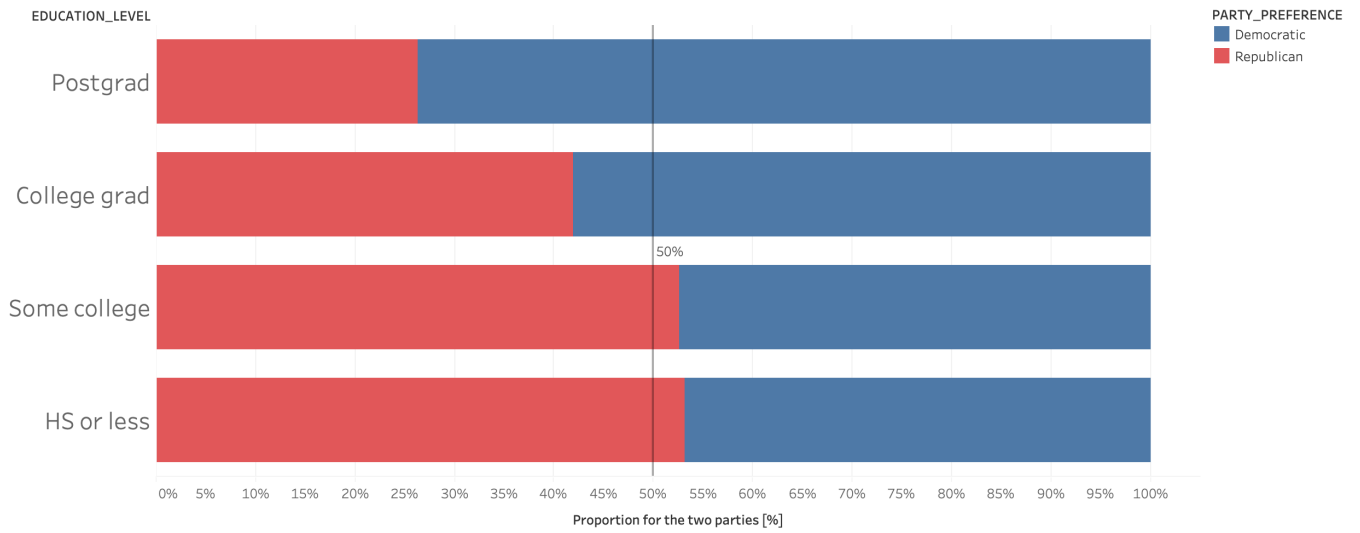


The trend of sum of Preferences for Party Preference. Color shows details about Party Preference. The marks are labeled by Education Level and sum of Preferences.

Displaying the preferences as % of Total computed by row (**Table (across)**)

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	SUM(PREFERENCES) (% of total by row)
Graph type:	Bar
Color:	PARTY_PREFERENCE
Size:	
Label:	

Bar Chart Preferences Proportion



% of Total PREFERENCES for each EDUCATION_LEVEL . Color shows details about PARTY_PREFERENCE . Percents are based on each row of the table.

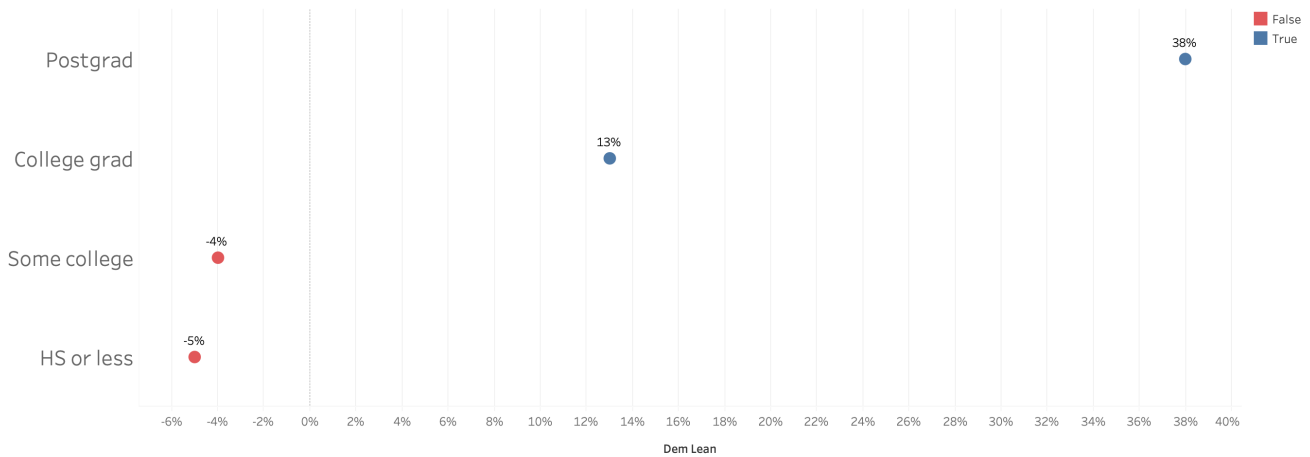
Another possibility is to show a derived measure: Lean of Dem over Rep (i.e. the difference in percentage points) that can be computed using the DemIndicator derived measure

Field	Dim./Measure	Formula
DemIndicator	Measure	<code>iif(PARTY_PREFERENCE=="Democratic",1,0)</code>
DemLean	Measure	<code>sum(PREFERENCES[DemIndicator])-sum(PREFERENCES(1-[DemIndicator]))</code>

The graph is then

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	DemLean
Graph type:	Circle
Color:	PARTY_PREFERENCE
Size:	
Label:	DemLean

Dot plot Dem Lean



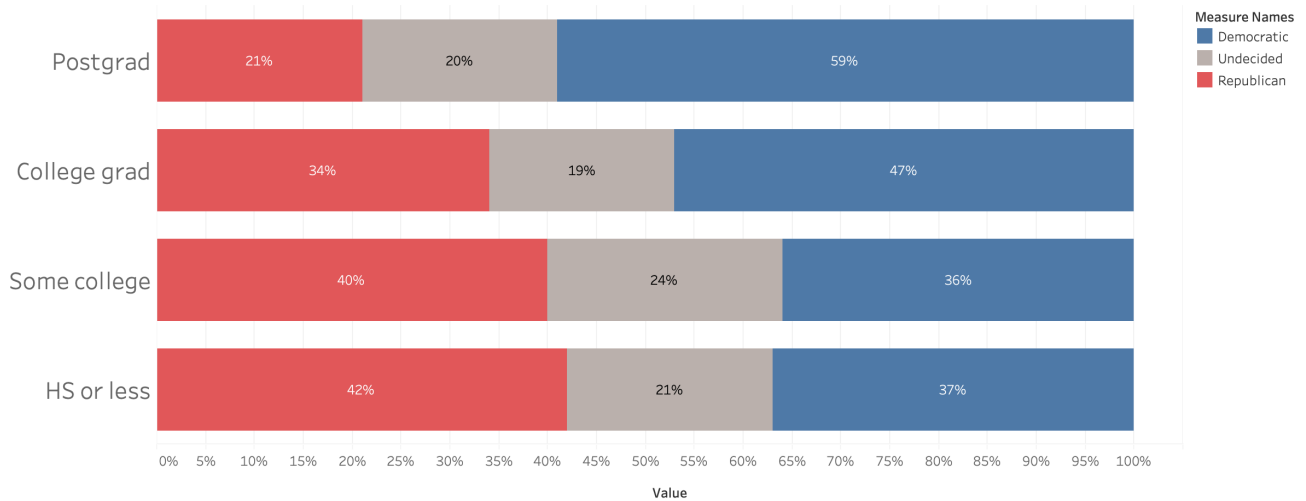
Dem Lean for each EDUCATION_LEVEL . Color shows details about FavorDem. The marks are labeled by Dem Lean.

we can compute the amount of "Undecided" as

Field	Dim./Measure	Formula
DemPref	Measure	$\text{DemIndicator} * [\text{Preferences}]$
RepPref	Measure	$(1 - \text{DemIndicator}) * [\text{Preferences}]$
Undecided	Measure	$1 - \text{sum}(\text{PREFERENCES}[\text{DemIndicator}]) - \text{sum}(\text{PREFERENCES}(1 - [\text{DemIndicator}]))$

Schema	Details
Rows:	EDUCATION_LEVEL
Columns:	MeasureValues: sum(DemPref) , Undecided, sum(RepPref)
Graph type:	Bar
Color:	MeasureNames
Size:	
Label:	

Bar Chart Preferences Stacked w/Undecided



Democratic, Undecided and Republican for each EDUCATION_LEVEL . Color shows details about Democratic, Undecided and Republican. The marks are labeled by Democratic, Undecided and Republican.