

DO YOU MISS YOUR DATA?

ONCE IN THE WHILE YOU WILL HAVE TO WORK WITH INCOMPLETE AND MISSING DATA. AVAILABLE VALUES CAN BE SPARSE ACROSS TIME AND SPACE. MOST OBVIOUS SOLUTION WOULD BE TO IGNORE THE GAPS AND WORK WITH WHAT YOU HAVE. HOWEVER, SOMETIMES THE ABSENCE OF DATA CAN LEAD TO INTERESTING CONCLUSIONS AND SHOULD BE INCLUDED IN YOUR VISUALIZATION. HERE ARE SOME ADVICES ON HOW TO APPROACH THIS ISSUE.

SHOW THE GAPS

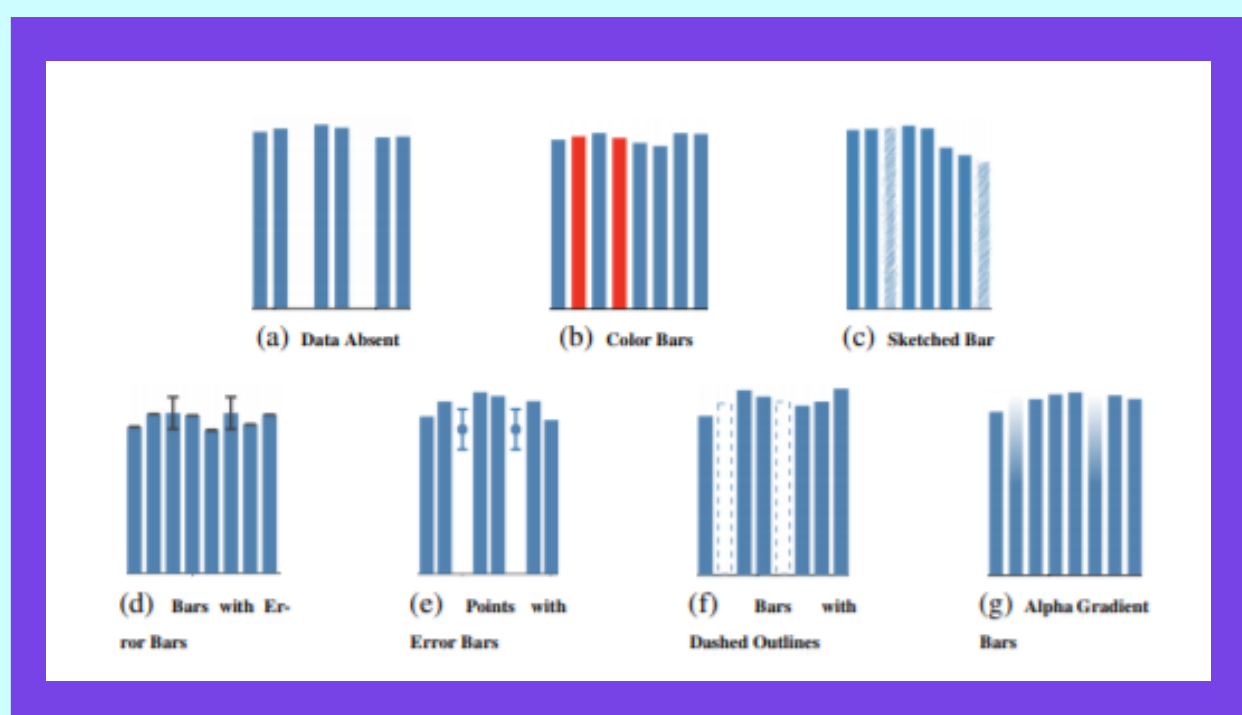
IF YOU MISS SOME DATA JUST SHOW IT.

IT MIGHT HAPPENS ESPECIALLY IN TIME SERIES THAT YOU MISS MONTH OR TWO.

STILL BETTER SHOW EMPTY SPACE (A, B, F) TO KEEP CONTINUITY THAT DELATE THIS MONTHS AT ALL. IT MAY LEAD A MISREPRESENTATIONS.

FROM THE OTHER SITE IF YOU MISS JUST SOME DAYS IT'S BETTER TO SHOW INSECURE THAN ASUME THAT WHAT YOU DON'T KNOW (D, E).

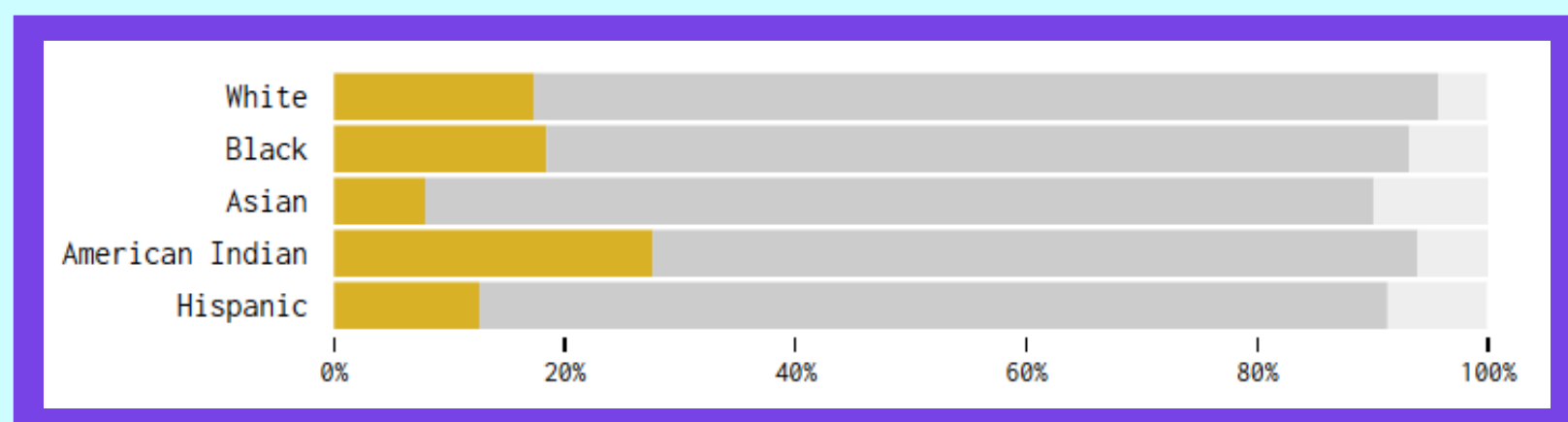
IF YOU ADDING UP, USE GRADIENT (G, C) TO PRESENT THAT IT'S NO ACCURATE .



IN SURVEY DATA OFTEN CONTAINS “NOT APPLICABLE” CATEGORY OR LEFT WITHOUT THE ANSWER AT ALL. IT'S OFTEN USEFUL TO TREAT THIS AS A SEPARATE ANSWER.

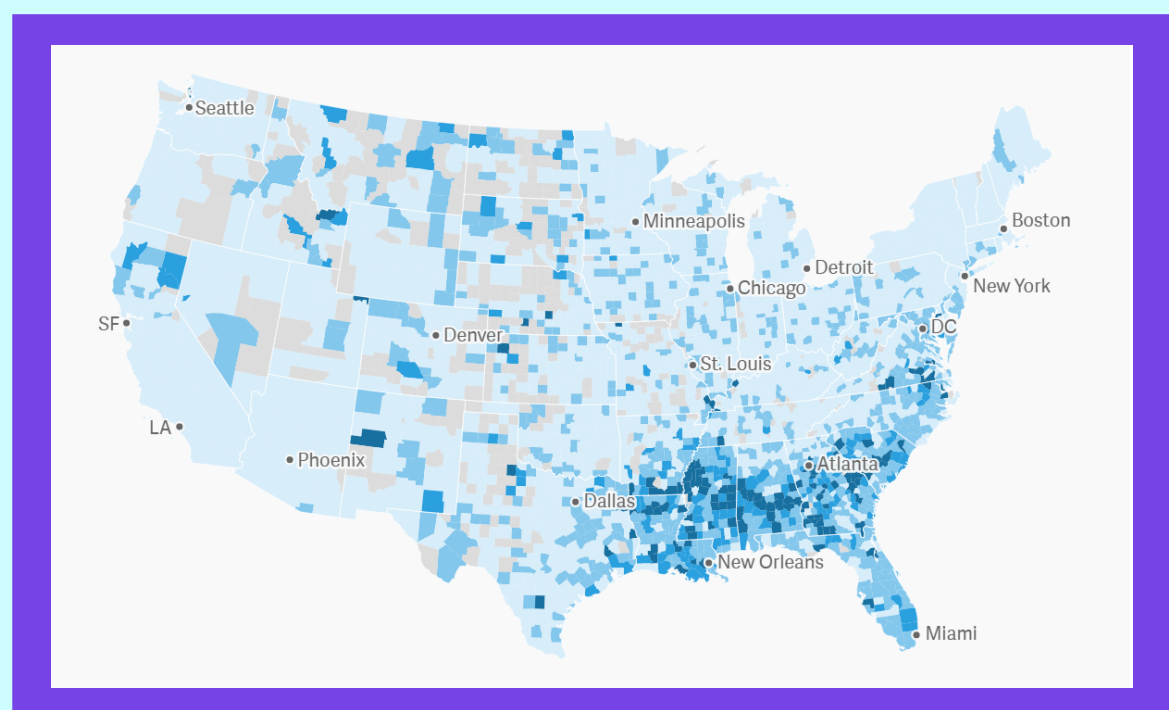
IF THERE IS NO BIG DIFERENCE BETWEEN YOU CAN CONNECT THEM. THEN TRY TO VISUALIZE WITH LEES STAND OUT COLOURS OR USE THIS DATA AS A SEPARATOR TO OTHERS ANSWERS.

THE CHART BELOW PRESENTS PERCENTAGE OF SMOKING PEOPLE (YELLOW) DEPENDS ON RACE AND ORIGIN. "NO RESPONSE" (LIGHT GREY) IS CONNECTED "LACK OF ANSWER" AND "DON'T WANT TO ANSWER" BECAUSE IT MAKE NO DIFFERENCE IN THIS PROBLEM.



ZOOM OUT

WHEN HANDLING MISSING DATA TRY TO ADJUST SIZE OF PRESENTED DATA SET SO THAT DEPENDENCIES BETWEEN REST OF THE DATA IS VISIBLE, WHILE VISUALIZATION REMAINS DETAILED. EXAMPLE BELOW PRESENTS HOW FREQUENT USAGE OF WORD “DEM” IS IN DIFFERENT STATES. ALTHOUGH THERE IS LACK OF INFORMATION ABOUT SPECIFIC STATES, USER CAN EASILY SEE GENERAL TREND. AT THE SAME TIME IT WOULD BE QUITE A CHALLENGE IF VISUALIZATION WOULD SHOW ONLY LEFT HALF OF A MAP.



INTERPOLATE

WHEN WORKING WITH IRREGULAR DATA THE MOST STATISTICAL APPROACH IS TO FILL IN THE GAPS IN A SMART WAY. FOR EXAMPLE, WE CAN ASSIGN TO MISSING VALUE WEIGHTED AVERAGE OF ADJACENT POINTS OR APPLY SIMPLE INTERPOLATION. WITH TIME SERIES REGRESSION MODEL CAN BE APPLIED SO THAT DEPENDENCIES CAN BE SEEN EVEN WITH MISSING DATA. DO NOT CHOOSE METHOD OF FILLING MISSING VALUES RANDOMLY BECAUSE IT COULD LEAD TO MAKING UNINFORMED CONCLUSIONS. SMOOTH CURVE OVER THE DATA IN EXAMPLE BELOW WAS MADE USING LOCAL REGRESSION.

