Golden Gate Park Access & Safety Program Survey

The comprehensive SFMTA survey database was obtained through a Sunshine Ordinance request. Many of the flaws and inconsistencies I will describe are baked into the structure of the survey itself. Others become rather obvious by just sorting and filtering the responses by different characteristics. Anybody with moderate spreadsheet skills could perform the same analysis.

Results of the Survey

City agencies, local journalists, and activist groups have promoted the apparently favorable response to the survey. Some have inaccurately claimed that 10,000 people responded to the survey and that the overwhelming majority of San Franciscans support a car-free JFK Drive. First of all, 10,000 people did not respond to the survey. There are 9,749 responses in the SFMTA database, and responses are not the same as people. As detailed below, there are many duplicate responses among the surveys submitted. The exact number of duplicates cannot be determined by anyone. However, there are certainly several hundred and potentially a couple thousand duplicate responses in the SFMTA database. Additional details are provided below. As a result, there may have only been 7,000 or fewer people that actually responded to the survey. In addition, not every survey respondent answered every question. For example, more than 900 responses did not answer the car-free JFK Drive question.

Second, the results of the survey cannot be extrapolated to the general population of San Francisco. You may be able to claim that a certain percentage of the survey responses favored car-free JFK Drive, but you cannot extend this claim to all of San Francisco. Responses to this survey were voluntary and anonymous. Only those who were aware of the survey and interested enough in the proposed changes in Golden Gate Park would have responded. As detailed later, many people responded to the survey multiple times. There was no attempt by SFMTA to ensure that the composition of responses reflected the actual demographic characteristics of San Francisco. Based on the demographic information that was provided on the surveys, the responses were clearly not representative of the general population of San Francisco. Non-white ethnic groups were significantly underrepresented. Many geographic areas of San Francisco were also significantly underrepresented. These issues are explained in greater detail below.

Advocates for a car-free JFK Drive have boasted that all income groups and ethnicities supported car-free JFK, based on the survey responses. However, they did not explain that the strongest apparent support came from high-income households, particularly those reporting incomes $200,000 and above. Responses from households earning $200,000 or more accounted for 30% of the total responses that provided income information on the survey. Advocates also failed to mention that ethnic diversity among the responses was severely lacking. Significantly inadequate representation by certain ethnicities in the survey is explained below in greater detail. Further, among all ethnicities, apparent support for a car-free JFK was strongest among White responses, which were heavily overrepresented in the survey. For example, there were nearly 1,200 survey responses from the following demographic group: White only, age 25-44, with household income of $150,000 or more. More than 93% of this demographic reportedly favored a car-free JFK Drive.

Alleged support for a car-free JFK steadily declined in older age groups. Only 38% of responses from residents age 65 and above favored the existing car-free JFK option (among nearly 1,300 responses from these age groups).

Survey responses that reported one or more disabilities also did not favor a car-free JFK Drive. Nearly 1,000 responses came from residents who reported one or more disabilities (the most common were mobility issues). Among those that answered the car-free JFK Drive question, only 38% allegedly favored the existing car-free JFK option.

Inadequate Representation from Certain Ethnicities

Nearly 7,000 surveys provided ethnicity information. The remaining surveys left these questions blank or stated that they preferred not to answer. Of the 7,000 surveys with ethnic information, 66% reported that they were White only (i.e. they did not select Hispanic or any other ethnic category other than White). According to U.S. Census information, about 40% of the population of San Francisco is comprised of non-Hispanic White residents. While White residents were significantly overrepresented in the survey, virtually all other remaining ethnicities were significantly underrepresented relative to their actual proportions of total city residents. For example, Asian & Pacific Islander responses represented less than 16% of the surveys, but these groups account for 35% of the total San Francisco population. Similarly, Hispanic & Latinx responses (of any race) comprised 8% of the survey responses, but these groups make up 15% of the total city population. The scenario is the same for Black & African American responses.

Inadequate Representation from Certain Geographies

There have been public claims by advocates that support for a car-free JFK Drive is strong citywide, with survey responses from 24 of 25 San Francisco zip codes desiring a car-free JFK. This claim, while perhaps technically accurate, is highly misleading, because there were very few survey responses from many San Francisco zip codes. Survey responses were recorded from 34 valid San Francisco zip codes. Seven of these zip codes represent P.O Boxes or other zip codes with no reported residents. Of the remaining 27 valid San Francisco zip codes with residents, 16 had fewer than 200 responses to the car-free JFK Drive question. For example, the zip code for Treasure Island (94130) had one response to the car-free JFK question. Zip code 94111 had 12 responses, and zip code 94129 had 25 responses to the car-free JFK Drive question. Support for a car-free JFK Drive cannot be reasonably determined for thousands of residents in those zip codes based on so few responses.

The southeastern part of San Francisco (zip codes 94112, 94124, and 94134) was particularly underrepresented in the survey. These three zip codes include the Ingleside, Excelsior, Bayview, Visitacion Valley, and Portola neighborhoods. These zip codes had about 300 total survey responses to the car-free JFK Drive question. Based on U.S. Census information, these three zip codes have more than 150,000 residents. Claiming that citizens in all of these neighborhoods favor a car-free JFK Drive based on only 300 combined survey responses is extremely disingenuous.

Further, there are some discrepancies in the zip code data from the paper surveys. This problem is described in more detail below. However, the number of responses from two zip codes (94107 and 94110) are not accurate, because SFMTA employees improperly entered this data for some paper surveys that lacked this information.

Duplicate Online Surveys and Problems with "Over-Vote" Analysis

The SFMTA identification and analysis of multiple responses from identical IP addresses was included with the database they provided to me. Just to be clear, this was their analysis, not mine. They did not provide individual IP addresses in the database (personal information), but they did indicate how many responses came from duplicate IP addresses. SFMTA identified about 2,000 online responses from duplicate IP addresses and estimated that about 1,700 of these responses should be excluded from the results (i.e. "over-votes"). I believe this process was significantly flawed, but at least SFMTA acknowledged that there is a very large number of duplicate online survey responses in their database. Nonetheless, their methodology improperly excluded many legitimate responses, while failing to identify many more likely duplicates.

I will describe the SFMTA methodology to identify and exclude "over-votes" in more detail. First, SFMTA identified online responses from IP addresses that had two or more responses to the existing car-free JFK question on the survey (blank responses were not counted – not every response answered this question). Based on this analysis, SFMTA identified a total of 2,023 potential duplicate survey responses from 836 IP addresses. That means that out of the 8,037 total online surveys that answered the car-free JFK question, about 25% came from duplicate IP addresses. Most of these responses came from IP addresses that reportedly had two or three responses each. However, some IP addresses had ten or more responses, including one IP address that reportedly had 50 responses to this question.

It is important to note that multiple responses from the same IP address did not always answer the question the same way (e.g. one response from an IP address may have indicated the existing car-free JFK option was "Desirable" and another response from the same IP address may have found that option "Undesirable"). SFMTA calculated "over-votes" by only looking for identical responses from the same IP address. So if there were two "Desirable" responses to this question from one IP address, both were identified as "over-votes." Unique responses from the same IP address were not considered "over-votes." Based on this analysis, SFMTA identified 1,706 "over-votes" among the 2,023 potential duplicate responses for this question.

This is a very flawed methodology, and the process certainly excluded many legitimate survey responses. For example, if a single person actually submitted two or more identical online responses to this question, SFMTA excluded all of these responses as "over-votes." Using this approach, such an individual would not have had any of their responses counted by SFMTA. In this situation, it would be more appropriate to count one response from such an individual, and to exclude any additional responses as invalid.

More importantly, it is virtually certain that different people sometimes submitted responses to the survey using a shared device (e.g. public computer) or from different devices that shared the same IP address (e.g. people using the same internet service provider network). Based on the SFMTA approach,

most of these responses would have been considered as "over-votes" and would have been excluded from the overall results, even though they likely originated from different people.

Remember, there was reportedly one IP address that had 50 responses to the car-free JFK question. Of these, 35 responses found the existing car-free JFK option "Very Undesirable." However, seven other responses from this same IP address answered this question "Very Desirable." The remaining eight responses from this IP address were distributed among all of the remaining options, including two responses that had "No Opinion." One of the 50 responses from this IP address answered the question "Desirable." In its calculation of "over-votes," SFMTA only counted any duplicate answers from the same IP address. In this instance, SFMTA considered 49 of the 50 responses from this IP address to be "over-votes." The only response that was counted was the single unique response for "Desirable." All other responses from this IP address were excluded because they had two or more answers for each option. Based on the distribution of answers to this question from this IP address, these responses do not appear to have come from a single person or even a few different people. It appears much more likely that this was either a shared device or a shared IP address used by very many people. However, virtually all of these responses were excluded by SFMTA as "over-votes."

There is another major problem with the analysis of duplicates and "over-votes." It would have been very easy for a single individual to submit multiple online surveys from different devices with different IP addresses, or to use many other available methods to submit multiple surveys from fake IP addresses. For example, an individual could have submitted surveys from their personal cell phone, their work laptop, and their family tablet. These devices would likely each have a unique IP address and would not be apparent as duplicate responses to the SFMTA.

In addition, there is no way for anyone to know how many people submitted both online surveys and paper surveys. There were also major problems with the paper surveys, as described in detail below.

Duplicate Paper Surveys and Problems with Data Manipulation

Paper surveys only accounted for about 9% of the total number of responses in the SFMTA database. However, there were major problems with these surveys. There appears to be a large number of duplicate paper surveys, and it also appears that SFMTA employees inappropriately entered some survey data for questions that were not answered on the original surveys. I contacted SFMTA and provided them with my analysis. SFMTA refused to answer my questions or provide any additional information regarding these issues.

I am 100% certain that there is a large group of duplicate paper surveys in the database. I know this because I found many duplicate pairs of paper surveys by carefully reviewing the underlying data. There are numerous pairs of surveys that are identical in every single respect, including identical answers to every question, verbatim comments in every comment field, and identical demographic information (e.g. age range, income range, and ethnicity). The most conclusive proof came from the optional comment fields. I easily identified duplicates where surveys had identical lengthy comments in multiple comment fields. There is no possibility that two surveys would have the exact same responses to dozens of questions and verbatim comments in every field. I asked SFMTA to explain how the duplicates got into the database and to accurately identify all duplicate paper surveys. I did not ask them to

confirm whether or not there are duplicates, because I know that there are. I suspect that a large batch of paper surveys was inadvertently entered into the database twice by different SFMTA employees. I estimate that potentially 50-60% of the paper surveys consist of these duplicate pairs, meaning that about 25-30% of the responses from the paper surveys should be discarded. By the way, SFMTA responded to my additional request with "No responsive records exist." I am following up with SFMTA with an additional request to obtain confirmation of my claims.

I am also 100% certain that SFMTA employees improperly entered zip codes for paper surveys that did not originally contain this information. I asked SFMTA to simply confirm this, and to identify any other instances where employees may have improperly entered or changed survey responses. I discovered the zip code discrepancies using the same methods I used to identify the duplicate paper surveys. Again, based on my review, it appears that at least two different SFMTA employees entered the paper surveys into the database. It appears that if the zip code was left blank, one employee always entered 94107 and the other employee always entered 94110 for these surveys. As a result, paper surveys from these two zip codes supposedly represent nearly 60% of all paper survey responses, which would be virtually impossible. For reference, these two zip codes aren't even near Golden Gate Park and they only represent about 12% of the city's population (there are more than 30 San Francisco zip codes). Also, these two zip codes only made up 10% of the online survey responses, which is much closer to the expected proportion. Another red flag: every single paper survey supposedly had a zip code entered. There were no blank responses for zip codes on any of the paper surveys. This is highly improbable. There were no other questions on the survey (paper or online) that didn't have at least some blank entries among the surveys submitted. The fact is, people don't always answer every question on a survey. By the way, SFMTA responded to my additional request with "No responsive records exist." While I do not believe that there was any ill intent behind this practice, it is a significant weaknesses with the administration of the survey. Entering or changing answers on behalf of survey respondents is a huge problem, even if it didn't affect a large number of surveys or change the overall survey results. This practice illustrates a significant weakness with survey controls. It has the potential to distort some of the survey results, and it calls into question the integrity of other survey responses.

Further, there is at least one other problem with the paper surveys. These surveys generally asked the same questions as the online surveys. However, the paper surveys did not ask for information regarding disabilities. Therefore, the total number of surveys from people reporting one or more disabilities is certainly undercounted by some degree. If the proportions were similar to the online surveys, there could have been another 100 responses from the paper surveys that would have reported one or more disabilities.

Additional Problems with the Surveys

I know for a fact that more than 500 online surveys were submitted and accepted after the public deadline of 5pm on November 24, 2021. The SFMTA database includes the submission date and time for every online survey. There were 552 online surveys submitted after the deadline, including 400 that were submitted after 12pm on November 28, 2021. Nearly 90% of these very late surveys favored car-free JFK. That should be a red flag, and all surveys received after the deadline should be discarded. The survey was well publicized and was open to the general public for more than two months. Anyone who

wanted to respond to the survey was perfectly capable of doing so within the established parameters, as understood by everyone.

I also know for a fact that there were nearly 600 surveys with contradictory responses. These included surveys that said they wanted both the existing car-free JFK and the No Project option (285 responses answered either "Desirable" or "Very Desirable" to both options). The total also includes responses that did not want either option (299 responses that found both options either "Undesirable" or "Very Undesirable"). Again, this is pretty easy to verify by using simple sorting and filtering techniques. These contradictions cloud the results, because some of the surveys that claim to support the existing car-free JFK also indicated that they wanted the roads reopened to cars (i.e. pre-pandemic status). This problem resulted from the poor structure and wording of the survey questions. The analysis of these two questions on the survey should exclude any contradictory responses.

About Michael Cawthon:

I have worked for a bank regulatory agency as an examiner and an analyst for over 30 years. The analysis and opinions in this paper are entirely my own and in no way reflect the work or views of my employer or any other organization.

Golden Gate Park Access and Safety Program
Survey available online 9/21/2021 through 11/24/2021 (although late responses were apparently accepted through 11/29/2021)
Paper surveys could also be submitted (although these surveys lacked questions about disabilities)

9,749 | **Total responses**

All of the information on this sheet is based on the actual data in the SFMTA spreadsheet
No adjustments have been made for late responses, invalid duplicate responses, or inappropriate manipulation of responses by SFMTA

Responses	Source	Proportion of Total
8,874	Online	91.02%
875	Paper	8.98%
9,749	**Total**	**100.00%**

Responses	Status	Proportion of Total
8,940	Complete	91.70%
809	Partial	8.30%
9,749	**Total**	**100.00%**

Responses	Language	Proportion of Total
9,705	English	99.55%
44	Chinese	0.45%
9,749	**Total**	**100.00%**

Responses	Question	Response Rate	
9,254	Fulton Loading Zone	94.92%	
9,175	Park Shuttle Revamp	94.11%	
9,108	Improve Parking Garage	93.42%	
9,062	Taxi Stands in Park	92.95%	
9,127	Bikeshare in Park	93.62%	
9,094	Scooters in Park	93.28%	
8,829	Existing Car-Free Route	90.56%	Note: This question had 8,037 responses online and 792 responses on paper
8,789	JFK Access Loop	90.15%	
8,795	Car-Free West End	90.21%	
8,810	No Project	90.37%	

Responses	Age Range	Proportion of Subtotal
54	Under 18	0.65%
289	19-24	3.49%
2,072	25-34	25.05%
2,071	35-44	25.04%
1,406	45-54	17.00%
1,023	55-64	12.37%
956	65-74	11.56%
399	75 or over	4.82%
8,270	**Subtotal**	**100.00%**
335	Prefer not to answer	
1,144	BLANK	
9,749	Total	

Responses	Ethnicity	Proportion of Subtotal	
1,084	Asian / Pacific Islander - only	15.56%	
185	Black / African American - only	2.66%	
357	Hispanic / Latinx - only	5.12%	
71	Middle Eastern / North African - only	1.02%	
27	Native American - only	0.39%	
4,597	White - only	65.98%	
151	Another race or ethnicity - only	2.17%	
495	Two or more ethnicities	7.10%	
6,967	**Subtotal**	**100.00%**	Total number of responses that indicated one or more ethnicities
1,473	Prefer not to answer (net of 20 duplicates)		Note: there were 20 responses that selected "Prefer not to answer" but also selected one or more ethnicities
1,309	BLANK		
9,749	Total		

Responses	Household Income Range	Proportion of Subtotal	
251	Less than $24,999	4.04%	
429	$25,000 - $49,999	6.90%	
700	$50,000 - $74,999	11.26%	
826	$75,000 - $99,999	13.29%	
1,271	$100,000 - $149,999	20.45%	
869	$150,000 - $199,999	13.98%	
1,870	$200,000 or more	30.08%	
6,216	**Subtotal**	**100.00%**	Total number of responses that indicated an income range
2,202	Prefer not to answer		
1,331	BLANK		
9,749	Total		

Responses	Disabilities	Proportion of Total	
77	Blind or vision impairment	0.79%	Note: The paper surveys did not contain questions about disabilities
143	Deaf or hearing impairment	1.47%	
661	Mobility disability	6.78%	
96	Cognitive or mental disability	0.98%	
196	Another disability or health condition	2.01%	
1,173	Subtotal (gross)	12.03%	
209	LESS: Duplicates (two or more)	2.14%	
964	**Subtotal (net)**	**9.89%**	Total number of responses that indicated at least one disability (adjusted for multiple entries)
73	MEMO: responses with two or more	0.75%	Total number of responses that indicated more than one disability (included in net subtotal above)
2,676	None (net of 4 duplicates)	27.45%	Note: there were 4 responses that indicated "None" but also selected one or more other disabilities
760	Prefer not to answer (net of 15 duplicates)	7.80%	Note: there were 15 responses that indicated "Prefer not to answer" but also selected one or more other options
5,349	BLANK	54.87%	
9,749	**Total**	**100.00%**	

Responses	Zip Codes	Proportion of Total	
8,647	Valid San Francisco Zip Codes	88.70%	Responses were reported from 34 valid San Francisco zip codes

897	Zip Codes Outside SF or Invalid	9.20%
9,544	Subtotal	97.90%
205	BLANK	2.10%
9,749	**Total**	**100.00%**

Responses	Ten Most Common Zip Codes	Proportion of San Francisco	
1,188	ZIP Code 94117	13.74%	Adjacent to GGP (Cole Valley, Haight-Ashbury, NOPA, Lower Haight)
1,126	ZIP Code 94122	13.02%	Adjacent to GGP (Outer Sunset, Inner Sunset)
953	ZIP Code 94118	11.02%	Adjacent to GGP (Inner Richmond, Presidio Heights)
935	ZIP Code 94110	10.81%	Mission, Bernal Heights
897	ZIP Code 94121	10.37%	Adjacent to GGP (Outer Richmond, Sea Cliff)
517	ZIP Code 94114	5.98%	Castro, Dolores Heights, Noe Valley
438	ZIP Code 94107	5.07%	Potrero Hill, Dogpatch, Central Waterfront
383	ZIP Code 94116	4.43%	Outer Sunset, Parkside, Forest Hill
350	ZIP Code 94115	4.05%	Pacific Heights, Western Addition, Anza Vista
291	ZIP Code 94131	3.37%	Forest Knolls, Glen Park
7,078	**Subtotal (top ten)**	**81.85%**	

Responses	Ten Most Common Zip Codes	Proportion of San Francisco	
1,188	ZIP Code 94117	13.74%	Adjacent to GGP (Cole Valley, Haight-Ashbury, NOPA, Lower Haight)
1,126	ZIP Code 94122	13.02%	Adjacent to GGP (Outer Sunset, Inner Sunset)

Golden Gate Park Access and Safety Program
Analysis of Favorable Responses to Car-Free JFK Drive Option by Age/Income
Figures exclude blank responses and "Prefer not to answer" responses for any of the relevant fields

Responses to Car-Free JFK Drive Option By Age Range									
Age Range	**Under 19**	**19-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75+**	**Totals**
Desirable or Very Desirable	42	252	1,842	1,706	975	581	381	108	5,887
Total	54	287	2,041	2,048	1,386	1,000	922	375	8,113
Percent	77.78%	87.80%	90.25%	83.30%	70.35%	58.10%	41.32%	28.80%	72.56%

Responses to Car-Free JFK Drive Option By Income Range				
Income Range	**Under $100K**	**$100K-$150K**	**$150K-$200K**	**Over $200K**	**Total**
Desirable or Very Desirable	1,426	929	664	1,596	4,615
Total	2,152	1,252	858	1,853	6,115
Percent	66.26%	74.20%	77.39%	86.13%	75.47%

Total Number of Responses to Car-Free JFK Drive Option									
Age/Income	**Under 19**	**19-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75+**	**Totals**
Under $100,000	20	151	573	349	249	280	357	147	2,126
$100,000-$149,999	2	37	389	283	217	162	120	37	1,247
$150,000-$199,999	4	17	253	240	152	96	72	24	858
$200,000+	2	26	489	686	385	156	74	30	1,848
Totals	**28**	**231**	**1,704**	**1,558**	**1,003**	**694**	**623**	**238**	**6,079**

Number of Responses to Car-Free JFK Drive Option - Answer "Desirable" or "Very Desirable"									
Age/Income	**Under 19**	**19-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75+**	**Totals**
Under $100,000	17	131	479	267	155	152	165	47	1,413
$100,000-$149,999	-	33	347	229	154	98	55	11	927
$150,000-$199,999	4	17	233	207	110	52	35	6	664
$200,000+	2	22	470	617	311	122	37	14	1,595
Totals	**23**	**203**	**1,529**	**1,320**	**730**	**424**	**292**	**78**	**4,599**

Percentage of Responses to Car-Free JFK Drive Option - Answer "Desirable" or "Very Desirable"									
Age/Income	**Under 19**	**19-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75+**	**Totals**
Under $100,000	85.00%	86.75%	83.60%	76.50%	62.25%	54.29%	46.22%	31.97%	66.46%
$100,000-$149,999	0.00%	89.19%	89.20%	80.92%	70.97%	60.49%	45.83%	29.73%	74.34%
$150,000-$199,999	100.00%	100.00%	92.09%	86.25%	72.37%	54.17%	48.61%	25.00%	77.39%
$200,000+	100.00%	84.62%	96.11%	89.94%	80.78%	78.21%	50.00%	46.67%	86.31%
Totals	**82.14%**	**87.88%**	**89.73%**	**84.72%**	**72.78%**	**61.10%**	**46.87%**	**32.77%**	**75.65%**