

How does CCP's Crackdown of Anti-Extradition Bill Movement Impact Hong Kong's Online Opinion?

Team Runaway

Computational Social Science Program

ABSTRACT

This study investigates the dynamics of online opinion in Hong Kong following the Crackdown of the Anti-Extradition Bill Movement and the implementation of the National Security Law in 2020. It employs text analysis of discussions from Hong Kong's largest online forum, LIHKG, to explore changes in political discourse. The research focuses on identifying shifts in sentiment, topic frequency, and expression styles to assess the impact of self-censorship on online freedom of speech. By comparing data from before and after the law's enactment, the study offers insights into the dynamics of self-censorship and its implications for public discourse.

KEYWORDS

Self-Censorship, Hong Kong National Security Law, Online Freedom of Speech, Internet Forums, Political Discourse, LIHKG, Text Analysis, Public Opinion, Censorship Dynamics.

Proposal Information:

Team Runaway. 2024. How does CCP's Crackdown of Anti-Extradition Bill Movement Impact Hong Kong's Online Opinion?. In *Proceedings of Computer Science with Social Science Applications 2* (Course code 'MACS 30122'). MACS30122 Chicago, IL, USA, 12 pages.

1 INTRODUCTION

The 2019-2020 Hong Kong Anti-Extradition Bill Movement was a series of massive demonstrations against introducing a bill to extradite the Fugitive Offenders Ordinance and the intervention of the China Communist Party (CCP) in Hong Kong's domestic political system. The movement caught global attention but ended bitterly in 2020 with the CCP's promulgation of the Hong Kong National Security Law on 30 June 2020. The law claimed to safeguard national security within the Hong Kong Special Administrative Region and deal with acts of secession, subversion of state power, terrorist activities, and collusion with foreign or offshore forces that endanger national security. Based on this law, the Hong Kong government prosecuted many dissidents and political activists who were previously engaged in the movement. In 2021, the CCP restructured the Hong Kong electoral system by introducing the National People's Congress Standing Committee's amendments to Hong Kong Basic Law, significantly undercutting prospects of universal suffrage in Hong Kong's future elections. With the new

amendments, only 88 out of the 470 seats in the Hong Kong Legislative Council are directly voted by Hong Kong citizens, with the rest controlled by the CCP to varying degrees. The turnout rate of the latest 2023 Hong Kong local elections dropped by 44% compared with the 2019 election, plummeting to a historical low of 27%. It could be expected that the public sphere in Hong Kong has been dramatically transformed after the Anti-Extradition Bill Movement.

How might public opinion in Hong Kong have been changed after the authoritarian crackdown by the CCP? Studies on the effect of the states' repression on dissents is an inconclusive field (e.g., Lichbach, 1987; Opp and Roehl, 1990). Recent research in this field still shows mixed results. While some studies find that repression could radicalize backlash (e.g., Sullivan and Davenport, 2017), other research shows a deterrence effect to the challenges against the state authority (e.g., Deibert, 2015; Pan and Siegel, 2020). Some scholars classify the repressions' effect on dissents based on regime types. For example, Gupta, Singh, and Sprague (1993) argue that while in democratic nations state repression would provoke a higher level of demonstrations, it might significantly reduce political oppositions in nondemocratic societies. However, as a society with democratic institutions but living under the shadow of authoritarian control, Hong Kong fits neither the category of a fully democratic nor nondemocratic society neatly. The hybrid political system in Hong Kong makes it more complex to disentangle the state crackdown's effect on its public sphere.

Moreover, the political attitudes in authoritarian-controlled societies differ publicly and privately. Kuran (1997) claims that citizens' private preferences and public preferences toward a regime might differ dramatically. It could then be asked if in the case of Hong Kong, there is a gap between citizens' publicly claimed political opinion gathered through survey and more privately communicated political preferences on social media. If so, what might this gap look like? This project ask this question based on the fact that although the political system in Hong Kong has been altered, the Chinese government has not mandated the same internet Great Wall and censorship systems in Hong Kong, allowing space for relatively free political discussion. This project hypothesize that the political preferences collected from social media might show more dissents than survey data in Hong Kong after the CCP's crackdown.

2 RESEARCH GAP AND QUESTIONS

The specific research questions in this project are: How does public opinion on the Hong Kong government and conditions of the society differ before and after the Anti-Extradition Bill Movement? How does such public opinion gathered in survey and on online social media forums differ? To do so, this project collects public opinion data from a Hong Kong political platform and survey data

Permission to make digital or hard copies of part or all of this work for personal or educational use is granted. To copy otherwise, or to redistribute to servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Course code 'MACS 30122, March 6, 2024, Chicago, IL'

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

from the Hong Kong Public Opinion Research Institute and analyzes their differences using methods including sentimental analysis. We hypothesize that 1) there is a gap between citizens' publicly claimed public opinion gathered through survey and more privately communicated political preferences on social media platforms and specifically that 2) the discussions on social media shows more negative sentiments than in survey data.

This project offer two advantages over the existing literature. First, the Hong Kong case could show how the foreign authoritarian rule impacts public opinion in a free and democratic society. The Anti-Extradition Bill Movement is a transiting point in Hong Kong history, marking the CCP's hijack of its electoral and legal system. This particular case may uncover more nuances in studying the effect of state repression on public opinion. Second, our studies utilize social media posts and sentimental analysis to survey public opinion. Few current studies use large-scale social media posts to analyze offline crackdown's impact on online public opinion, with Pan and Siegel in 2020 claiming that they are the first scholars to do so. Moreover, in the case of Hong Kong, where censorship at large allows for free political discussion, social media posts might better capture the average citizen's political opinion. Our comparison between social media data and survey data could comprehensively represent Hong Kong citizens' political orientations after the crackdown.

3 DATA COLLECTION

This project employs two data sources: LIHKG, a Cantonese political platform in Hong Kong, and the public opinion survey data from Hong Kong Public Opinion Research Institute. The data collected in LIHKG are posts and replies from this forum, focusing on the difference in the discussion related to political issues before and after the implementation of the HKNSL as the research object. LIHKG is the largest online forum in Hong Kong, with nearly 20 million monthly visits, surpassing the combined popularity of Hong Kong Twitter and Reddit. It features a dedicated political discussion section. During the Anti-Extradition Law Amendment Bill Movement in 2019, LIHKG played a crucial role as a platform for participant interaction and information dissemination, and it was targeted by DDoS attacks from Mainland China. Survey data indicated that 33% of respondents actively participated in discussions on the forum by replying to posts related to the movement, while 10.8% contributed by creating new posts (Ku, 2020).

Given the absence of LIHKG datasets online, this project implemented a Python-based web scraping program to collect pertinent data. Data processing of political opinions needs to consider the impact of unexpected, influential events on political discourse. For instance, let's say a government scandal broke out in October 2023; this could result in a spike in negative public opinion toward the government. However, these sentiments may be more reflective of the event's immediate influence rather than the broader political context at the time. Therefore, this project intends to use a comprehensive political ideology scale involving key terms as the foundation for data collection. This scale, devised by Erni & Zhang (2022), contains terms relevant to the three main themes prevalent in online Hong Kong political discourse: "Democracy and Justice," "Protest and Resistance," and "Government and Governance." The

program leverages the site's search function to locate posts and replies containing these keywords, facilitating subsequent scraping activities.

民主与公义 (Democracy and Justice)	抗议与抗争 (Protest and Resistance)	175
民主 (Democracy)	示威者 (Protesters)	181
公义 (Justice)	示威 (Protest)	182
公平 (Fairness)	抗争 (Resistance)	183
投票 (Voting)	文宣 (Propaganda)	184
利益 (Interest)	打壓 (Suppression)	185
保衛 (Defense)	对抗 (Confrontation)	186
普選 (General Election)	落街 (Hit the Streets)	187
政府与治理 (Government and Governance)		188
政府 (Government)	特首 (Chief Executive)	189
黑社会 (Triads)	滲透 (Infiltration)	190
極權 (Authoritarianism)		191
		192

Table 1: Classification source: Erni, J. N., & Zhang, Y. (2022). *Wild hopes: Sourcing the political vocabulary of digital citizenship from the LIHKG Forum*. International Communication Gazette, 84(4), 349–375. <https://doi.org/10.1177/17480485221094123>

Due to LIHKG's prominent role in Hong Kong's social movements, it experienced a DDoS attack from mainland China. The forum is particularly sensitive to bot activity and unusual access requests. The design of the web pages involves dynamic loading, meaning the content within posts and replies can only be fully viewed after scrolling on the page post-load; a static scraping method fails to retrieve all the information about the posts. Additionally, this website also has an anti-web scraping mechanism based on IP addresses that blacklists and blocks an IP if it frequently accesses different threads and posts on the site. Consequently, this project faced two significant challenges: the need for the scraping tool to handle dynamically loaded websites to access hidden information that requires scrolling, and the ability to bypass anti-scraping mechanisms that detect bots through user behavior analysis.

To address the dynamic loading, the project used Selenium, which can simulate interactions between humans and web browsers. Selenium, via WebDriver, can execute typical user actions like clicking, scrolling, and entering data—actions that help mimic a real user's browsing pattern, thus allowing the project's scraper to circumvent detection mechanisms. In addition, Selenium's capability to execute JavaScript for dynamic content loading and access information through simulated user actions like scrolling ensured that the scraper could collect complete data.

To circumvent the site's IP-based anti-scraping measures, the scraping program incorporated a rotate IP strategy. Rotate IP tools enable the scraper to change its IP address throughout the scraping process. These tools work similarly to a VPN, rerouting internet traffic through a proxy server to the website, which then sees the proxy server's IP rather than the actual IP during visits. This approach, by distributing requests across different IPs, significantly reduces the chances of being blocked.

233 4 DATA ANALYSIS

234 In total, this project scrapped 57,368 posts and replies. In the data
 235 wrangling process, this project use Pandas to structure the scraped
 236 data into CSV format, including each post and reply's user ID,
 237 the content, and time. To segment Chinese (Cantonese) words, this
 238 project uses a tool called Pycantonese to split the content of a mes-
 239 sage into meaningful individual words or phrases for later process-
 240 ing. When needed, this project will use the Microsoft Translator
 241 API to translate Cantonese to English automatically, as it is one
 242 of the few tools that support Cantonese to English translation and
 243 provide an API, while neither Google Translate nor DeepL offers
 244 translation services to Cantonese.

245 For the scraped data, this project first did the exploratory data
 246 analysis by getting top tokens over the years and their frequency
 247 trend to see the changes of the popular topics. Next, this project
 248 used the TensorFlow framework to fine-tune a pre-trained BERT
 249 model for sentiment analysis. Since the data used in this project
 250 consisted of Chinese and forum-like texts, the model utilized a
 251 dataset of 10k posts from Weibo (the largest social media platform
 252 in mainland China) labeled with a total of six emotions to help the
 253 model identify sentimental expressions in Chinese texts.

254 When processing the survey data, this project downloaded the
 255 CSV files related to the keywords from the website and translated
 256 them into relevant line graphs. Each chart shows the views of Hong
 257 Kong people on a particular issue or project over a specific period.

258 Through sentiment analysis of text data, this project discovered
 259 that anger is the predominant sentiment expressed by Hong Kong
 260 residents when discussing political issues. Anger dominated the
 261 discussions across three different themes, accounting for the largest
 262 proportion of emotions, particularly in discussions categorized un-
 263 der "Government and Governance," where the proportion of anger
 264 even exceeded half (see Appendix, figures 4, 5, 6). This indicates
 265 that in the forum's political discussions, people tend to express
 266 their negative views more frequently while seldom expressing positive
 267 emotions. Trend analysis of discussions revealed that political
 268 dialogues related to the three themes peaked during the 2019 Hong
 269 Kong Anti-Extradition Bill protests. At this point, there was a gen-
 270 eral trend of expressing universal anger about political issues, yet
 271 it is noteworthy that this anger coexisted with neutral discussions.
 272 This implies that protesters and netizens used online forums not
 273 just as an outlet for negative emotions but also as a space for ra-
 274 tional discourse on political issues. Despite a significant decline in
 275 discussions across all three themes after the social movements of
 276 2019, a notable observation is that discussions related to "protest
 277 and resistance" almost completely vanished post-2019 (see Appen-
 278 dix, figure 2), aligning with the actual lack of major social move-
 279 ments in Hong Kong that year. This suggests that the chilling ef-
 280 fect induced by the National Security Law not only diminished ac-
 281 tual social movements but also stifled the organization and discus-
 282 sion of such movements on the Internet, further reducing the like-
 283 lihood of protests. Discussions related to "Democracy and Justice,"
 284 however, showed seasonal fluctuations, spiking towards the end
 285 of each year (see Appendix, figure 1), maintaining a balance of an-
 286 gry and neutral sentiments, possibly due to Hong Kong's district
 287 council elections occurring during this period. The discussions on

291 democracy and a just governance system seemed not to have de-
 292 clined as dramatically as those on social movements, still showing
 293 seasonal fluctuations linked to the electoral cycle. It is essential
 294 to note the exceptionally pronounced surge in angry discourse at
 295 the end of 2020 regarding "Democracy and Justice," possibly pro-
 296 voked by the mainland government's enactment of the National
 297 Security Law, sparking public discontent. Discussions on "Govern-
 298 ment and Governance" have consistently been dense, which might
 299 imply that fear of the National Security Law may be more confined
 300 to concerns about being criminally charged for participating in or
 301 stopping social movements, but this fear and silence have not gen-
 302 eralized across all discussions on government-related topics.

303 Observations from survey data show that Hong Kong citizens' 304 attitudes towards the government and society turned negative dur-
 305 ing the Anti-Extradition movement and positive afterward, align-
 306 ing with levels before the movement (see Appendix, figures 7, 8, 9). 307 This trend is broadly consistent with forum data. Still, it seems not 308 to capture the periodic fluctuations and detailed categorizations 309 corresponding to forum discussions, suggesting that while survey 310 data might explain trends in public opinion, they may not cap- 311 ture the nuances and specific categories of change. Additionally, 312 current forum data seems to present more politically charged dis- 313 cussions with angry sentiments than before the Anti-Extradition 314 movement, which is not reflected in survey data. The surveys even 315 suggest that after 2022, public trust and affirmation of democracy's 316 development exceeded levels before the movement. The discrep- 317 arity between survey and forum data could potentially be due to 318 the influx of mainland immigrants post-movement, who would 319 have equal opportunity to be sampled in surveys and might poten- 320 tially offer more positive views of the government. Yet, they are 321 unlikely to familiarize themselves with and use Hong Kong's pol- 322 itical forums immediately after immigrating, potentially leading 323 to the observed divergence in forum data, which may stem from 324 differences in government and political discourse between native 325 Hongkongers and immigrants. Overall, we discussed our hypoth- 326 esis that suppression of the anti-extradition movement and enact- 327 ment of the national security law might have led to a chilling effect 328 on internet discourse and that there could be gaps between public 329 opinion gathered from surveys and online social media platforms. 330

331 5 CONCLUSION

332 In summary, this project has developed two main tools. The first 333 is a dynamic web scraping tool based on Selenium, which can also 334 perform IP rotation with the help of additional tools. This allows 335 it to correctly scrape web pages that require dynamic loading and 336 evade website IP address blocking. The second tool is a fine-tuned 337 BERT model built using the TensorFlow framework. This model 338 was trained on a dataset of 10k Weibo posts (the largest social 339 media platform in Mainland China), annotated with six different 340 emotions. It is capable of identifying expressions of emotion in Chi- 341 nese social media texts, enabling sentiment analysis of data from 342 Chinese forums. Further, using the collected data, the project 343 conducted sentiment analysis and visualization. It was observed that 344 in survey data, Hong Kong citizens' attitudes towards the Hong 345 Kong government and society became negative during the Anti- 346 Extradition movement and turned positive after the movement, 347

similar to pre-movement levels. However, network data indicated a significant deterioration in Hong Kong citizens' attitudes towards the government and society during the movement, with negative emotions like anger remaining prevalent in discussions post-movement. This change in discourse might be attributed to the chilling effect of the National Security Law, as Hong Kong netizens almost ceased discussing issues related to social movements following the suppression of the Anti-Extradition movement and the enactment of the law. There also appears to be a potential gap between public opinion gathered from surveys and online social media platforms.

6 GITHUB REPOSITORY

The code used for this project has been made available on GitHub at the following URL: <https://github.com/mac30122-winter24/final-project-runaway>

7 LIMITATION

The limitation of this project and the direction for future improvement is the better utilization of the data scraped and the results processed from sentiment analysis for more reliable modeling and statistical analysis. Due to the limited time for this project, there are many things that were initially planned but not successfully accomplished. Firstly, concerning data analysis, the report mainly presents descriptive statistics and exploratory analysis and lacks the establishment of sufficiently credible statistical models and the review and design for causal inference, rendering the interpretation of the data more exploratory and speculative than rigorous analysis. Additionally, the initial comparison between forum data and survey data also poses significant issues, as the project did not develop and design a methodology that effectively matches forum data with survey data. The keywords and themes involved in the collected survey data do not perfectly correspond to the data scraped from the forums, which considerably limits the credibility of comparisons between different data types.

8 ETHICS

This study involves web scraping and analysis of political discussions, raising issues of privacy and sensitivity of the subject matter. Firstly, regarding privacy and anonymity, forum users might not anticipate their posts being utilized for research purposes. Even though this information is publicly available, consent for research use is not implicitly granted by the users. To address this issue, the study will avoid collecting or recording personally identifiable information, such as usernames, to ensure that all data used in the research findings remain anonymous.

Secondly, due to the sensitive nature of political content, it must be handled with caution to avoid disseminating harmful or misleading information. Prior to commencing the research, the proposed content will be scrutinized to ensure that it does not violate any Institutional Review Board (IRB) guidelines. The study will also acknowledge that the models used for textual analysis involve a translation process, and information may become distorted during cross-language translation. Any assumptions, limitations, and potential biases in interpretation arising from this should be reported.

Furthermore, risks associated with data security and storage are present. Thus, restrictive measures will be implemented to prevent unauthorized access or data breaches, and secure data disposal will be conducted when it is no longer needed.

REFERENCES

- [1] Bar-Tal, D., *Self-Censorship as a Socio-Political-Psychological Phenomenon: Conception and Research*. Political Psychology, 38(S1), 37–65, 2017. <https://doi.org/10.1111/pops.12391>
- [2] Deibert, R., *Cyberspace under Siege*. Journal of Democracy, 26(3), 64–78, 2015. <https://doi.org/10.1353/jod.2015.0051>
- [3] Erni, J. N., & Zhang, Y. *Wild hopes: Sourcing the political vocabulary of digital citizenship from the LIHKG forum*. International Communication Gazette, 84(4), 349-375, 2022.
- [4] Gupta, D. K., Singh, H., & Sprague, T., *Government Coercion of Dissidents: Deterrence or Provocation?* Journal of Conflict Resolution, 37(2), 301–339, 1993. <https://doi.org/10.1177/0022002793037002004>
- [5] King, G., Pan, J., & Roberts, M. E., *How Censorship in China Allows Government Criticism but Silences Collective Expression*. American Political Science Review, 107(2), 326–343, 2013. <https://doi.org/10.1017/s0003055413000014>
- [6] King, G., Pan, J., & Roberts, M. E., *Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation*. Science, 345(6199), 2014. <https://doi.org/10.1126/science.1251722>
- [7] Ku, A. S., *New Forms of Youth Activism—Hong Kong’s Anti-Extradition Bill Movement in the Local-National-Global Nexus*. Space and Polity, 24(1), 111–117, 2020.
- [8] Kurian, T., *Private truths, public lies: The social consequences of preference falsification*. Harvard Univ. Press, sec. print, 1997.
- [9] Lee, J. L., *PyCantonese: Cantonese linguistic research in the age of big data*. Talk at the Childhood Bilingualism Research Centre, the Chinese University of Hong Kong, 2015.
- [10] Lichbach, M. I., *Deterrence or Escalation?: The Puzzle of Aggregate Studies of Repression and Dissent*. Journal of Conflict Resolution, 31(2), 266–297, 1987. <https://doi.org/10.1177/0022002787031002003>
- [11] Ong, E., *Online Repression and Self-Censorship: Evidence from Southeast Asia*. Government and Opposition, 56(1), 141–162, 2019. <https://doi.org/10.1017/gov.2019.18>
- [12] Opp, K.-D., & Roehl, W., *Repression, Micromobilization, and Political Protest*. Social Forces, 69(2), 521, 1990. <https://doi.org/10.2307/2579672>
- [13] Pan, J., & Siegel, A. A., *How Saudi Crackdowns Fail to Silence Online Dissent*. American Political Science Review, 114(1), 109–125, 2020. <https://doi.org/10.1017/S0003055419000650>
- [14] Sullivan, C. M., & Davenport, C., *The Rebel Alliance Strikes Back: Understanding the Politics of Backlash Mobilization*. Mobilization: An International Quarterly, 22(1), 39–56, 2017. <https://doi.org/10.17813/1086-671X-22-1-39>

APPENDIX

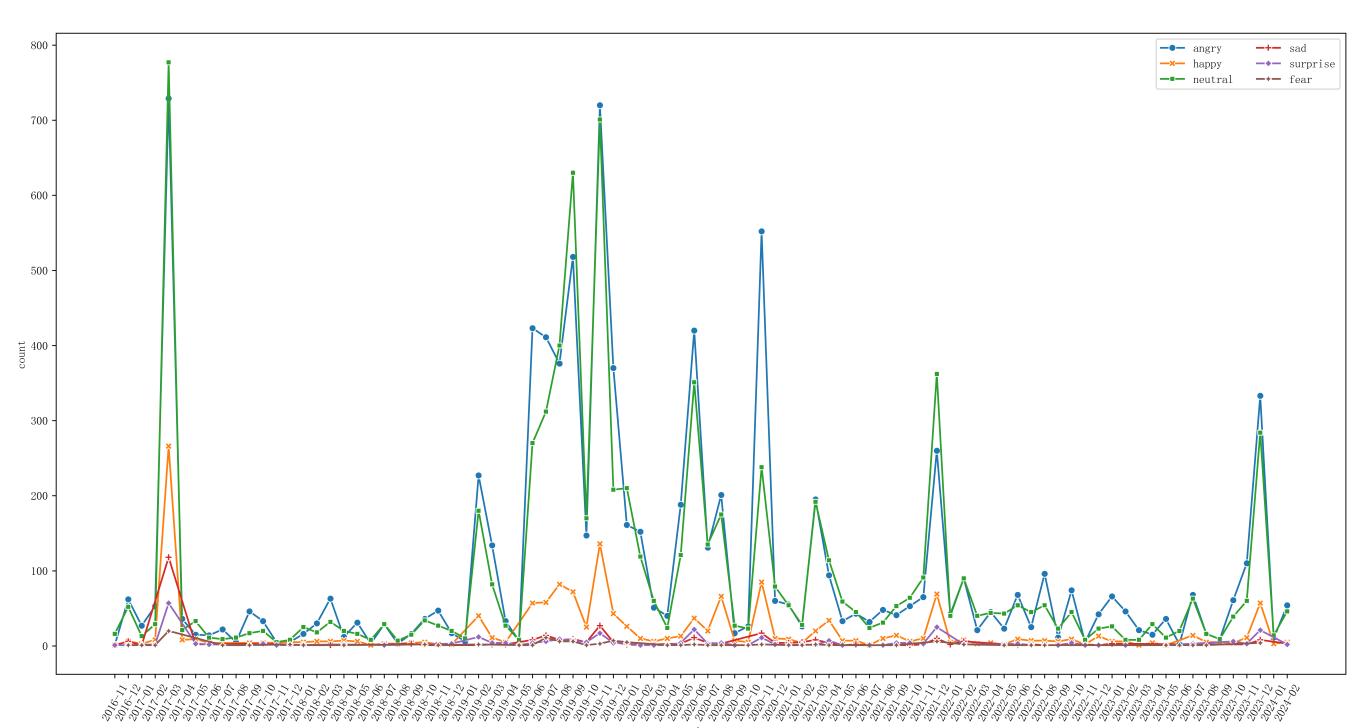


Figure 1: Sentimental trend for topic "Democracy and Justice"

523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580

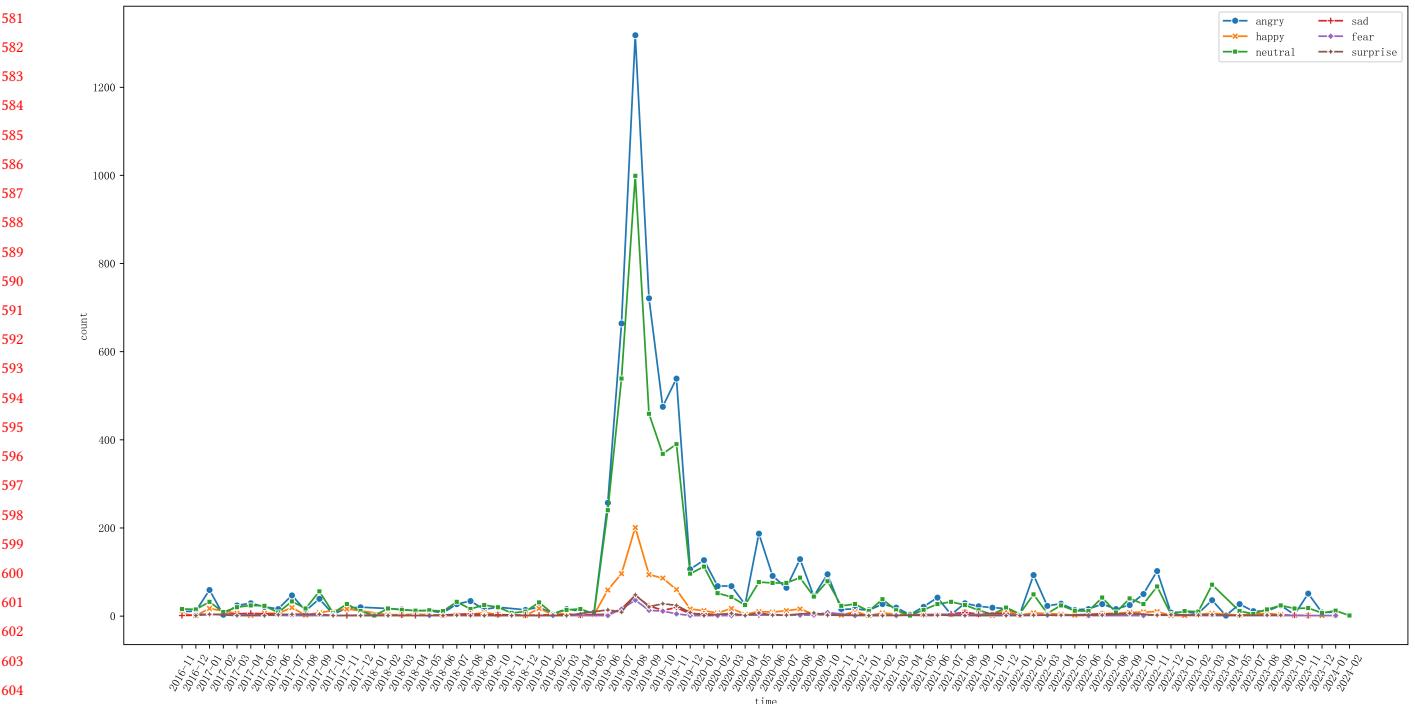


Figure 2: Sentimental trend for topic "Protest and Resistance"

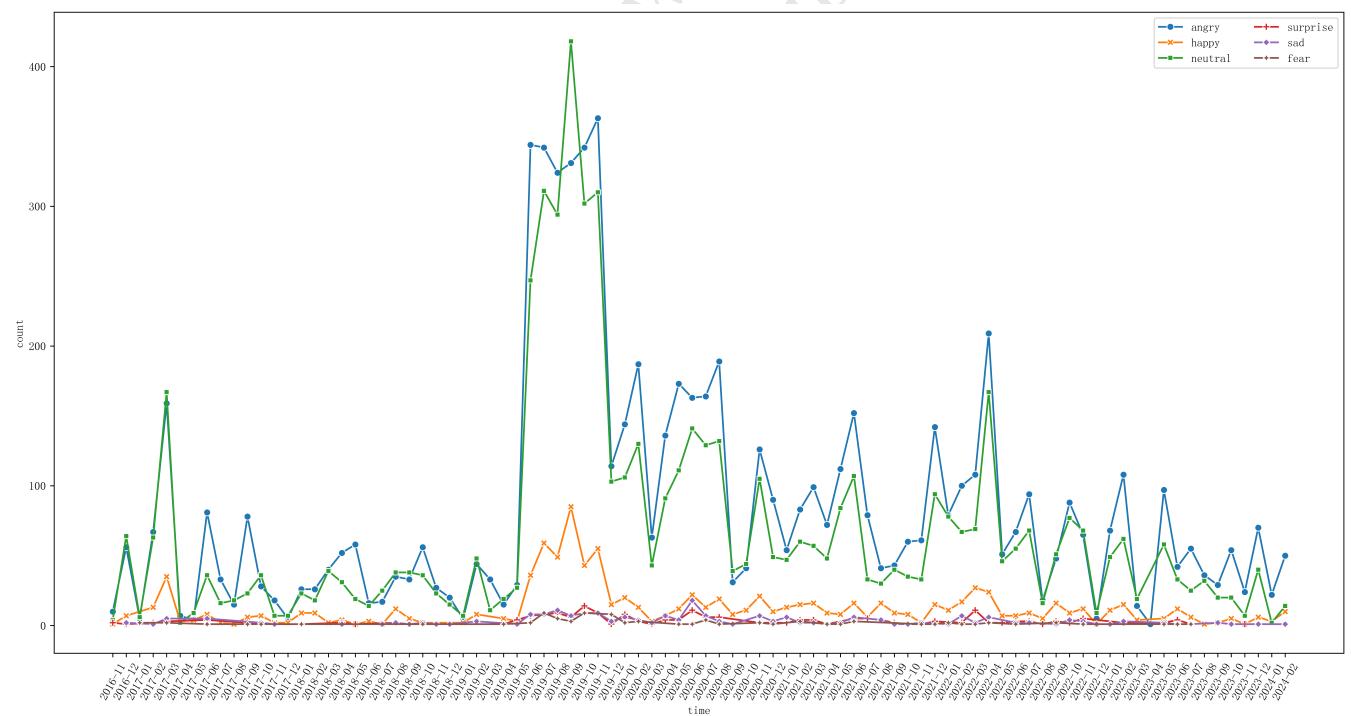


Figure 3: Sentimental dis for topic "Government and Governance"

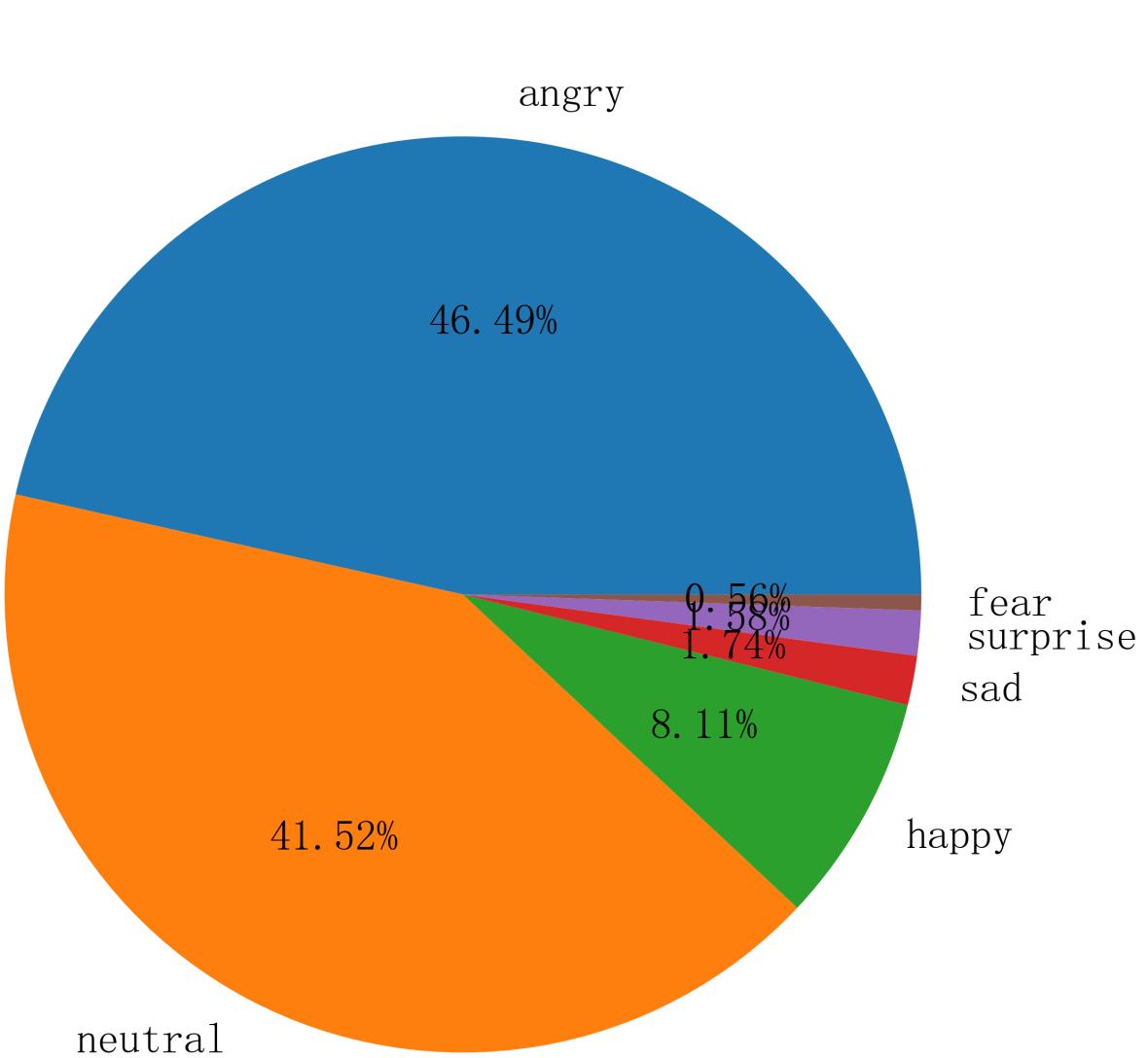


Figure 4: Sentimental distribution for topic "Democracy and Justice"

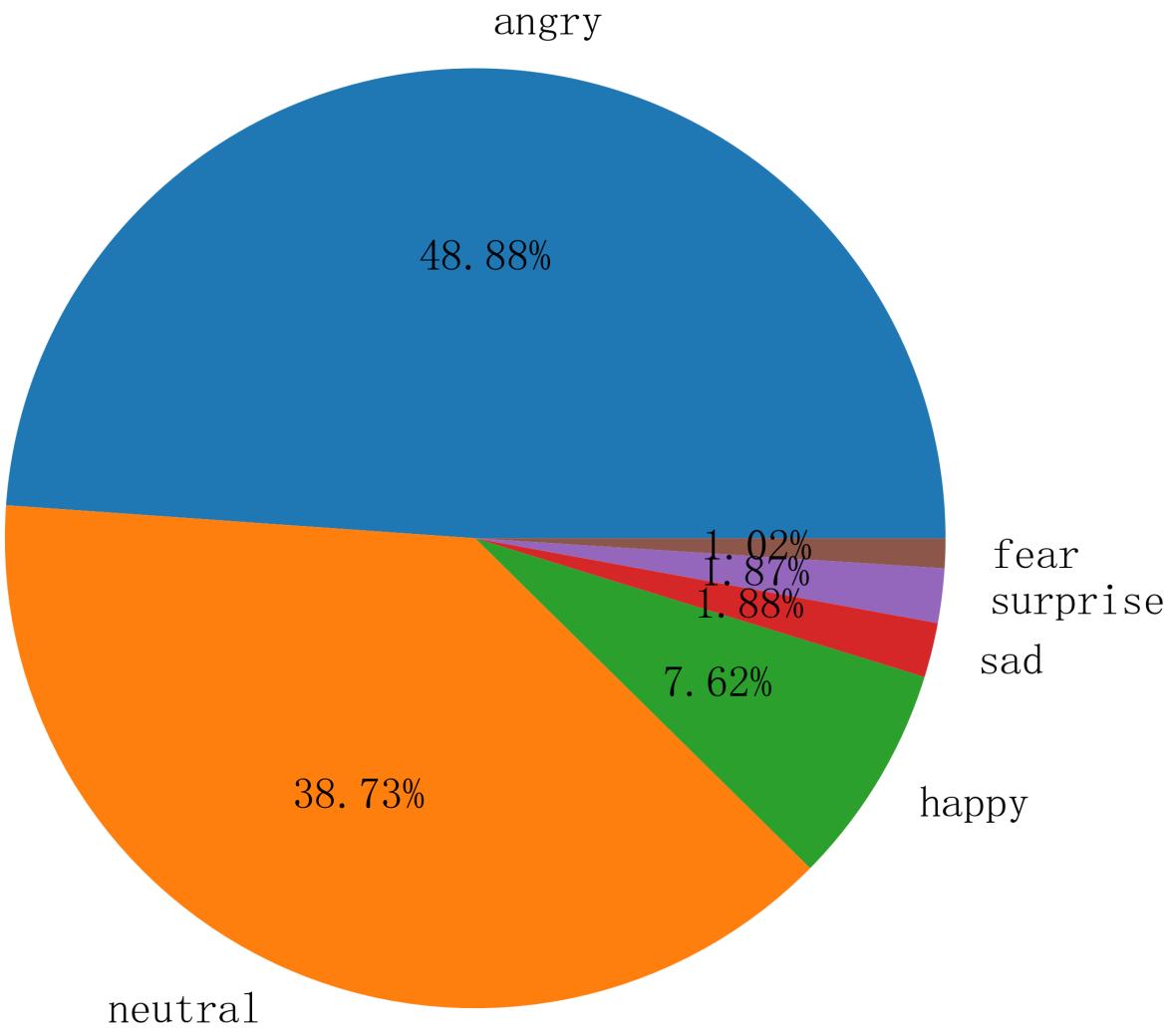


Figure 5: Sentimental distribution for topic "Protest and Resistance"

813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928

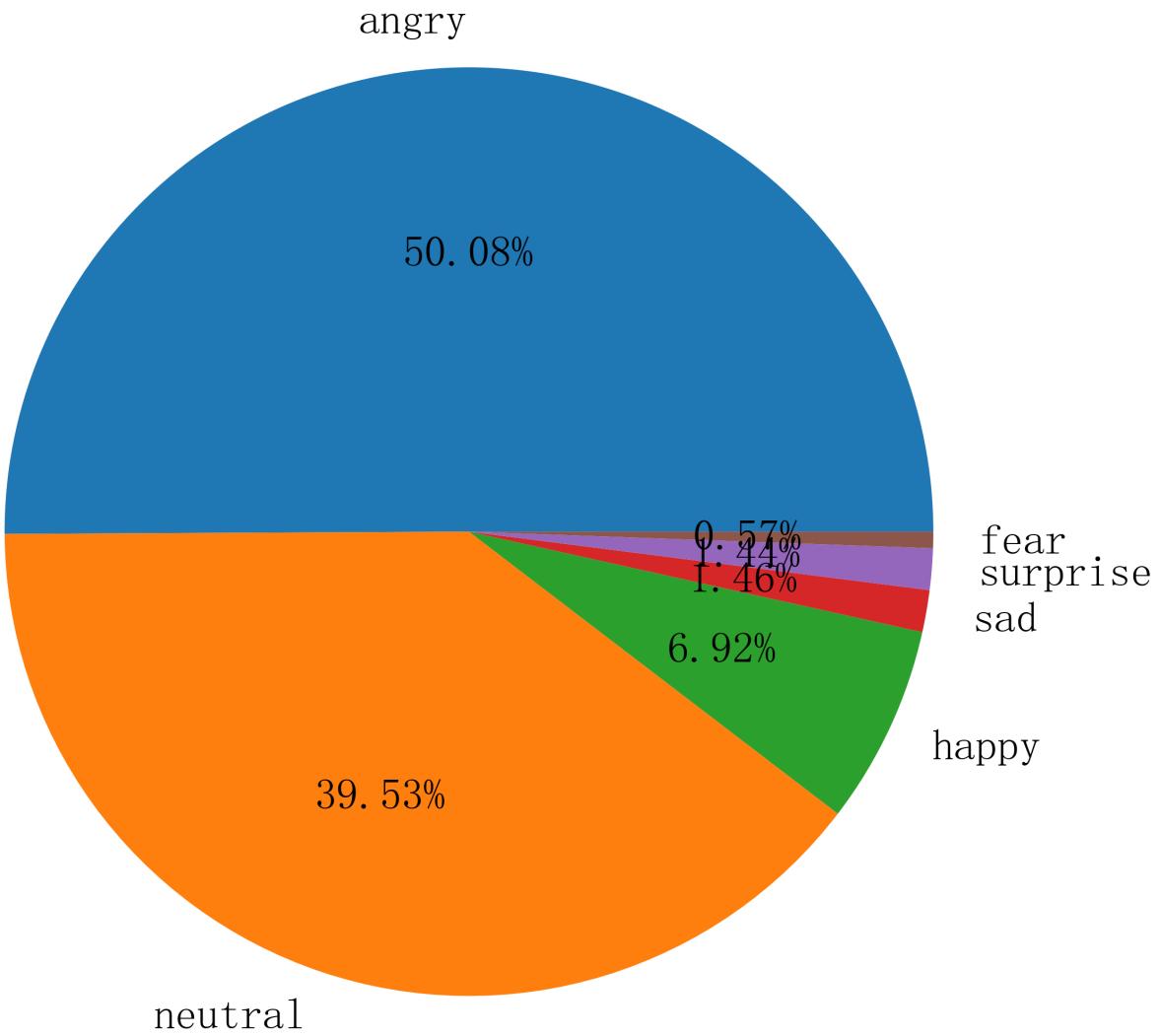


Figure 6: Sentimental distribution for topic "Government and Governance"

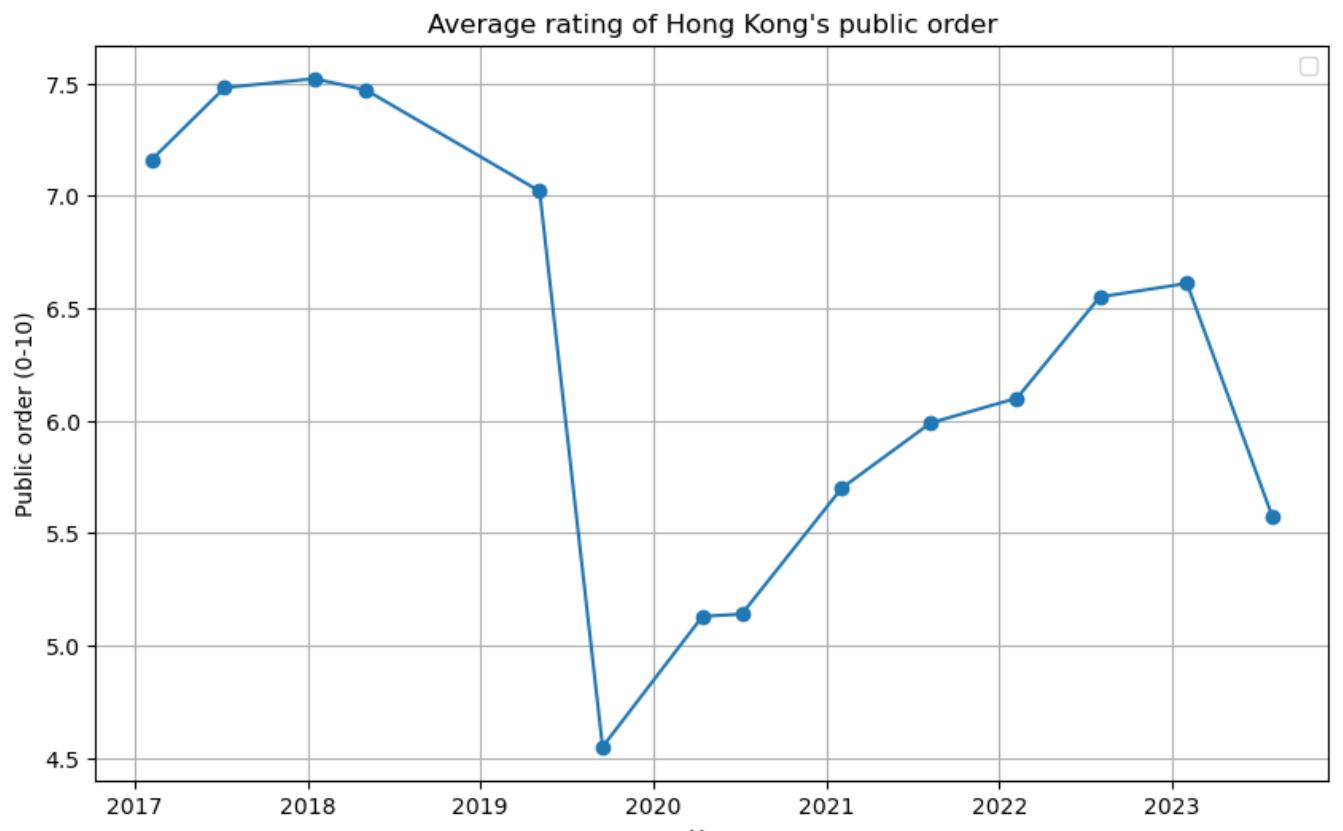


Figure 7: Survey data for public order

1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102

1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160

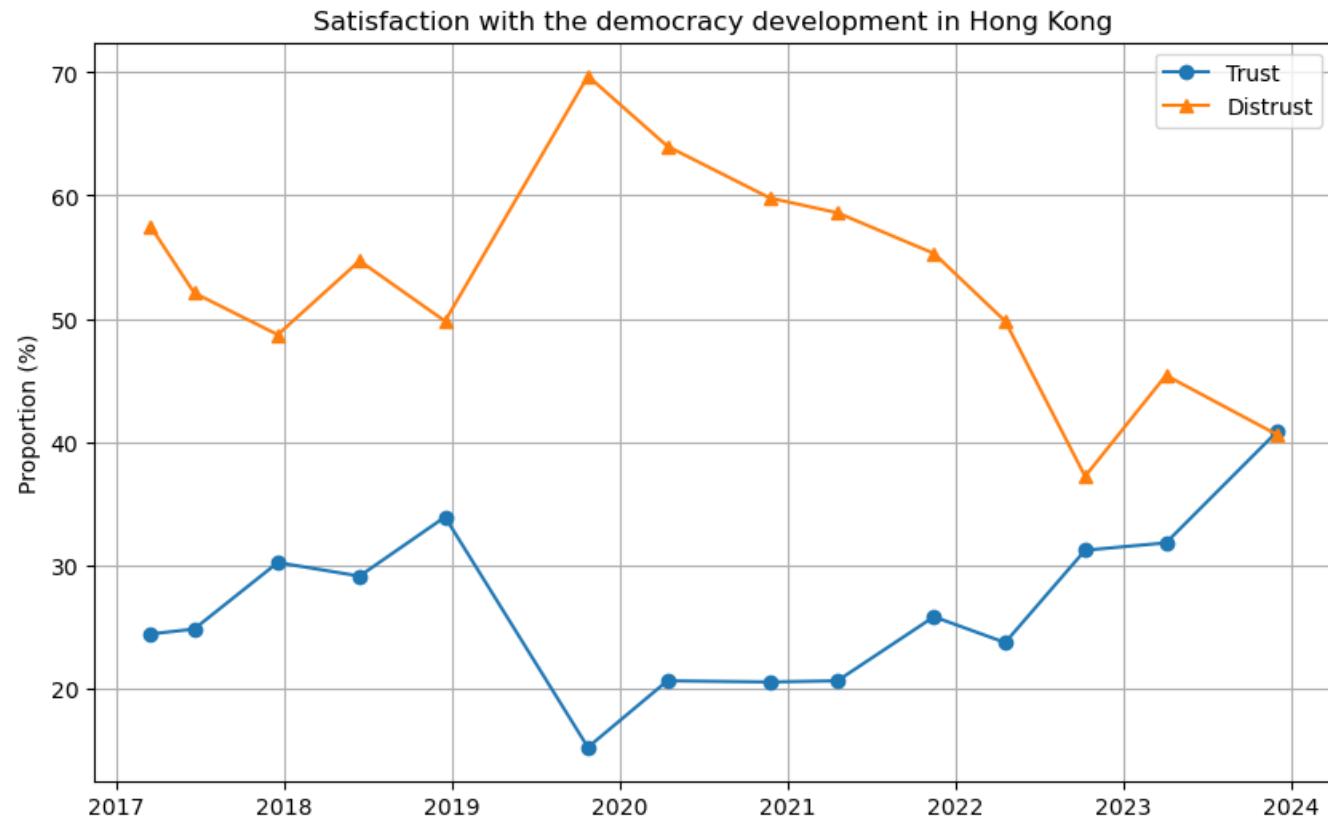


Figure 8: Survey data for democracy

1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276

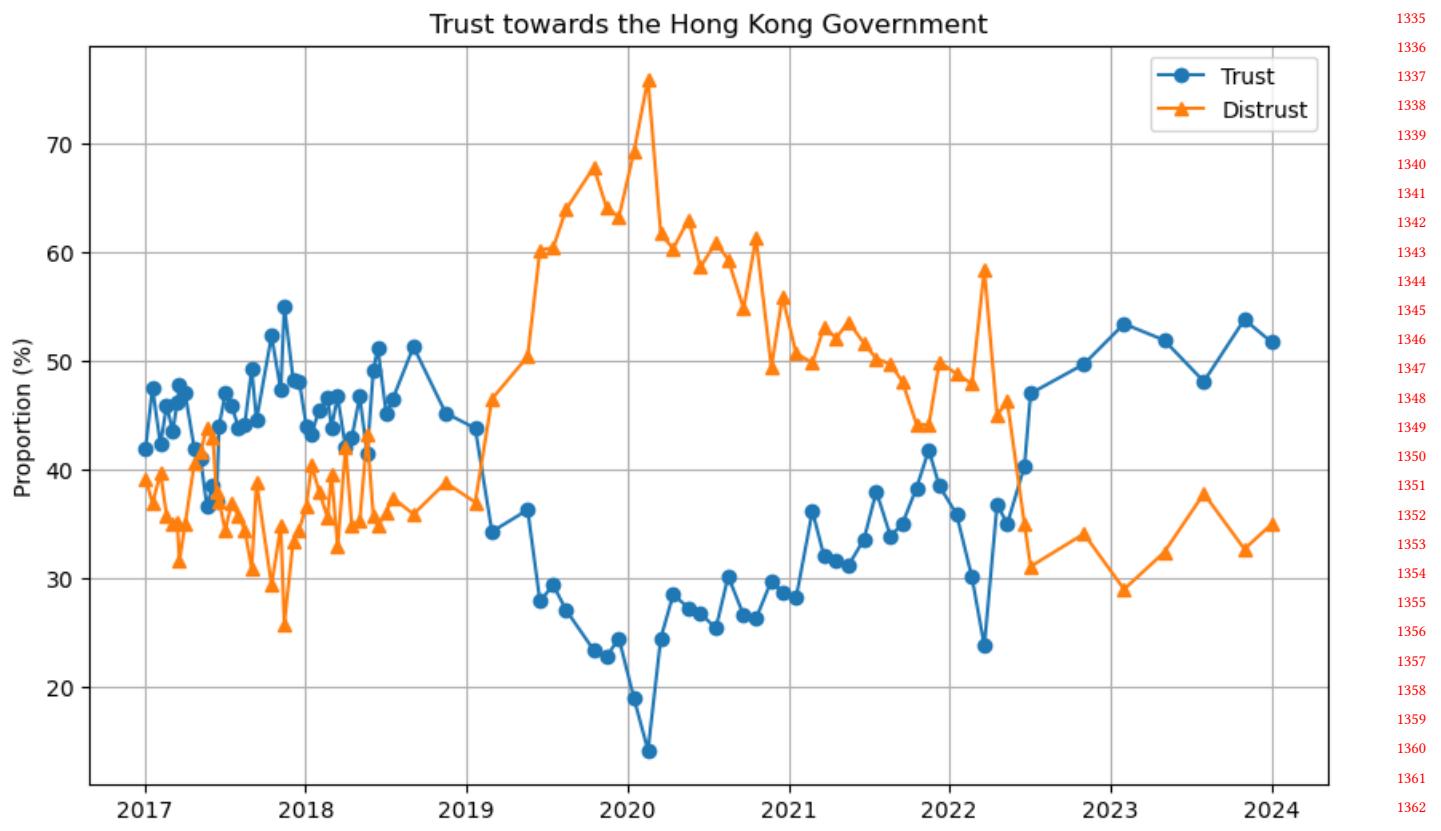


Figure 9: Survey data for trust in government