

The results on the datasets released under the names `clinicaltrial_2019.csv.gz` and `clinicaltrial_2020.csv.gz` are supplied in this document to allow you to check your code. These correspond to a very basic implementation: if you do any more sophisticated (pre-)processing, you will likely obtain different results – (assuming no bugs) this is fine, but you need to explain why the results are different in your report (you may find it easier to include basic results as well).

## Question 1

2019 326348  
2020 356466

## Question 2

	2019	2020
Interventional	255945	277631
Observational	64163	71434
Observational [Patient Registry]	6171	7332
Expanded Access	69	69

## Question 3

	2019		2020	
1	Carcinoma	11155	Carcinoma	12245
2	Diabetes Mellitus	9830	Diabetes Mellitus	10425
3	Neoplasms	7815	Neoplasms	8534
4	Breast Neoplasms	7486	Breast Neoplasms	8009
5	Syndrome	6842	Syndrome	7419

## Question 4

	2019		2020	
1	C04	123221	C04	133091
2	C23	113997	C23	124589
3	C14	82043	C01	94293
4	C10	76665	C14	88065
5	C01	73477	C10	83894
6	C06	73272	C06	79382
7	C08	45180	C08	61138
8	C13	36293	C13	39316
9	C18	36051	C18	38439
10	C12	34699	C12	37294

## Question 5

2019:

1. National Cancer Institute (NCI): 3003
2. Merck Sharp & Dohme Corp.: 2124
3. M.D. Anderson Cancer Center: 2097
4. Mayo Clinic: 1930
5. Novartis Pharmaceuticals: 1881
6. Assistance Publique - Hôpitaux de Paris: 1764
7. Massachusetts General Hospital: 1715
8. Hoffmann-La Roche: 1694
9. National Taiwan University Hospital: 1608
10. Eli Lilly and Company: 1500

2020:

1. National Cancer Institute (NCI): 3100
2. M.D. Anderson Cancer Center: 2238
3. Merck Sharp & Dohme Corp.: 2184
4. Mayo Clinic: 2097
5. Assistance Publique - Hôpitaux de Paris: 2043
6. Novartis Pharmaceuticals: 1962
7. Massachusetts General Hospital: 1823
8. Assiut University: 1806
9. Hoffmann-La Roche: 1761
10. National Taiwan University Hospital: 1720

## Question 6

	2019	2020
Jan	1368	1544
Feb	1149	1286
Mar	1470	1740
Apr	1368	1080
May	1342	1176
Jun	1647	1424
Jul	1547	1237
Aug	1406	1126
Sep	1421	1167
Oct	1310	1176
Nov	1223	1078
Dec	2690	2084