

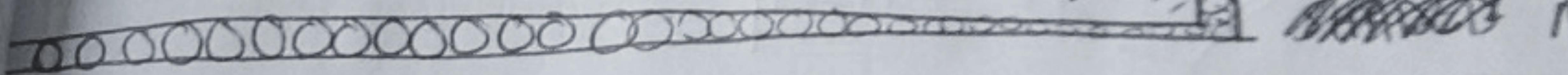
A promotional photograph of the cast of the television show "Freaks and Geeks". Seven teenagers are posed in front of a row of wooden lockers. From left to right: a boy in a dark leather jacket; a girl in a green jacket over a striped shirt; a boy in a white t-shirt; a boy in a plaid shirt; a boy in a blue and yellow striped polo shirt; a boy wearing glasses and smiling; and a boy in a plaid shirt. The boy in the center foreground has a surprised or shocked expression.

FREAKS AND GEEKS

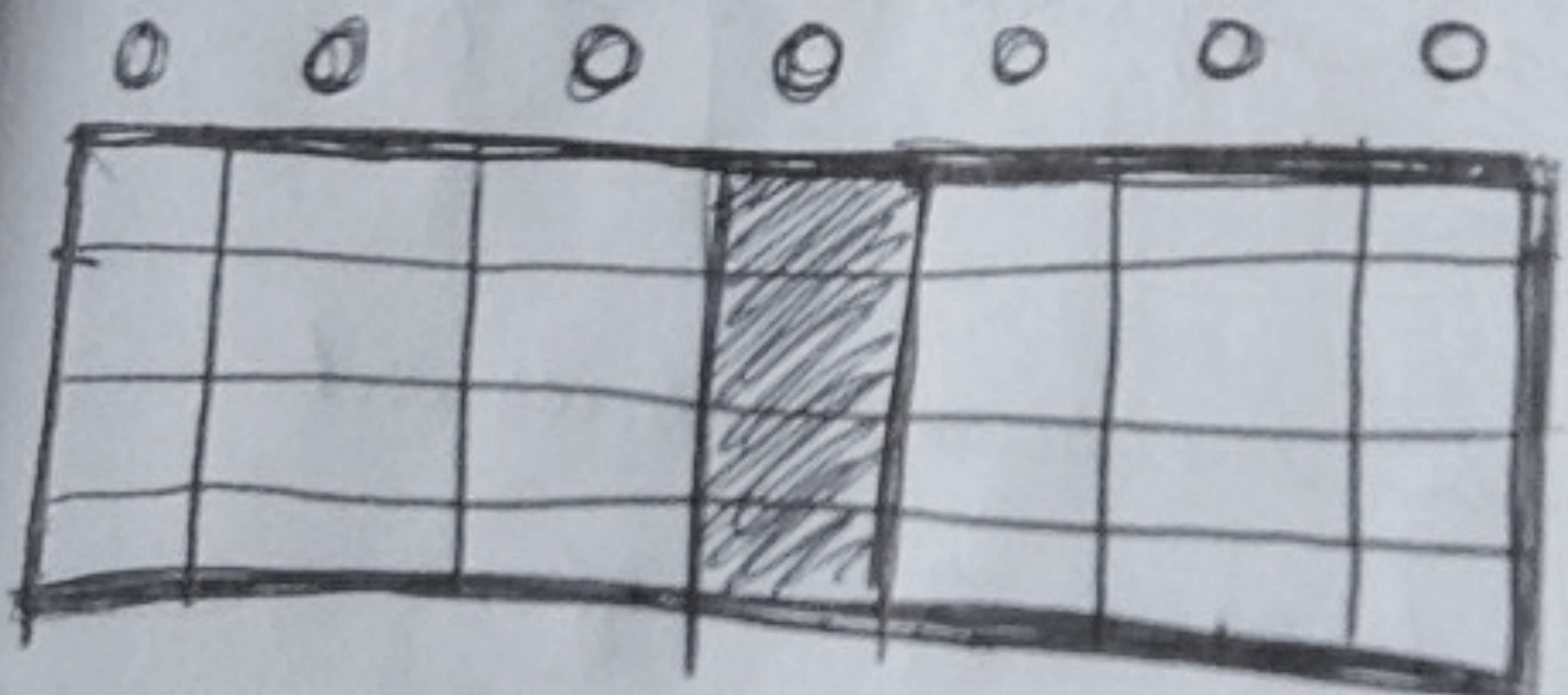
A group of six people are standing in front of a row of wooden lockers. From left to right: a woman wearing a dark green beret and a dark denim jacket; a man in a white t-shirt and a light blue denim vest; a woman with long brown hair wearing a dark green corduroy jacket over a striped shirt; a man in a dark t-shirt and a plaid shirt; a bald man in a blue and white striped polo shirt; and a man on the far right wearing glasses and a dark plaid shirt. The man in the blue and white striped polo has his mouth open as if speaking.

FREQSANDGEEKS

WHAT IS OUR PLAN!?



CIRCULAR
STRUCTURE



TOMORROW VS

What are these data about?

Who are we designing for?

30 HOUR PROCESS

CONCEPTS
-DQ
-USER
-GOALS

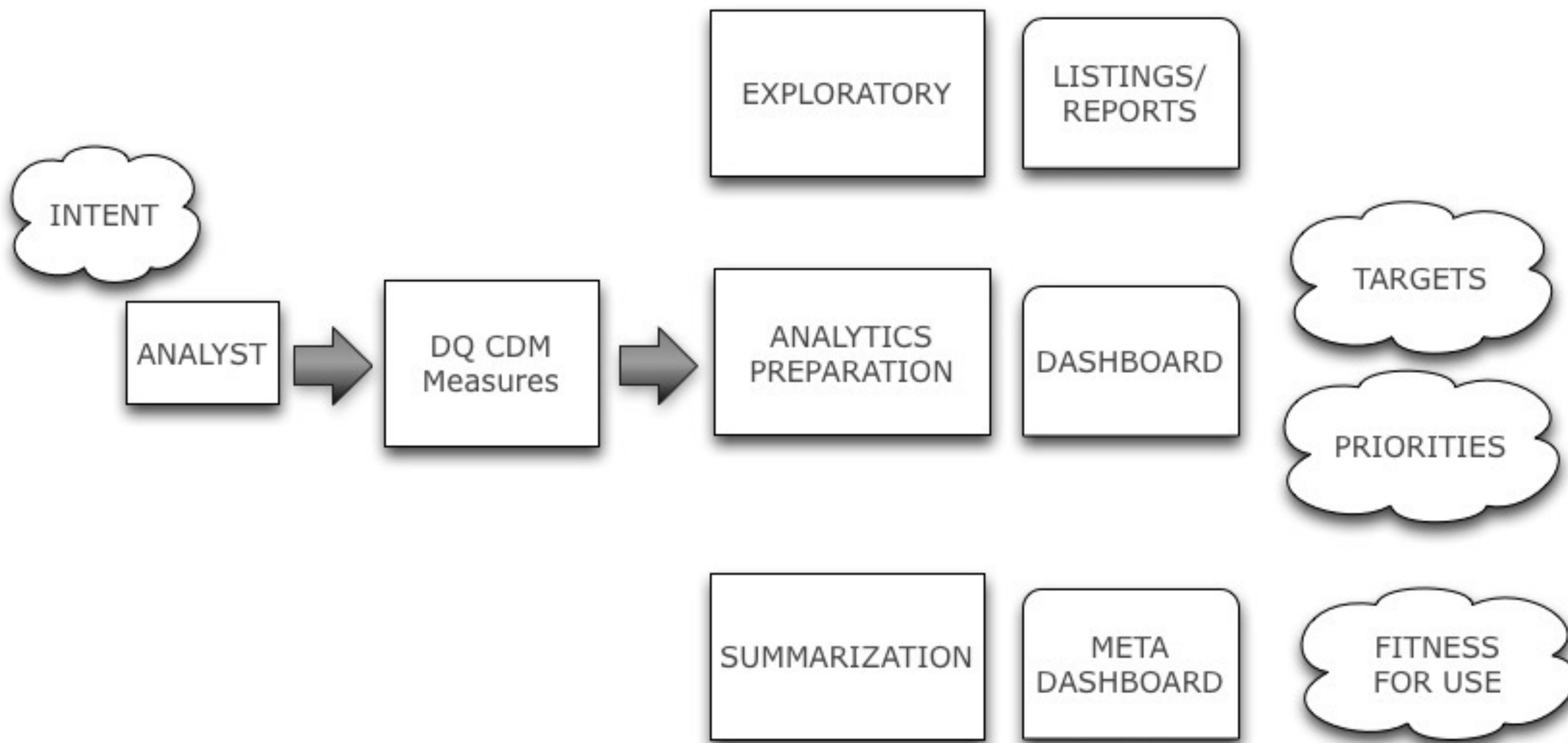
EXPLORE DATA
(R/SAS)

PROTOTYPING
(Illustrator/
InVision)

DIGEST DATA
(R/SAS)

VISUALIZATION
(D3)

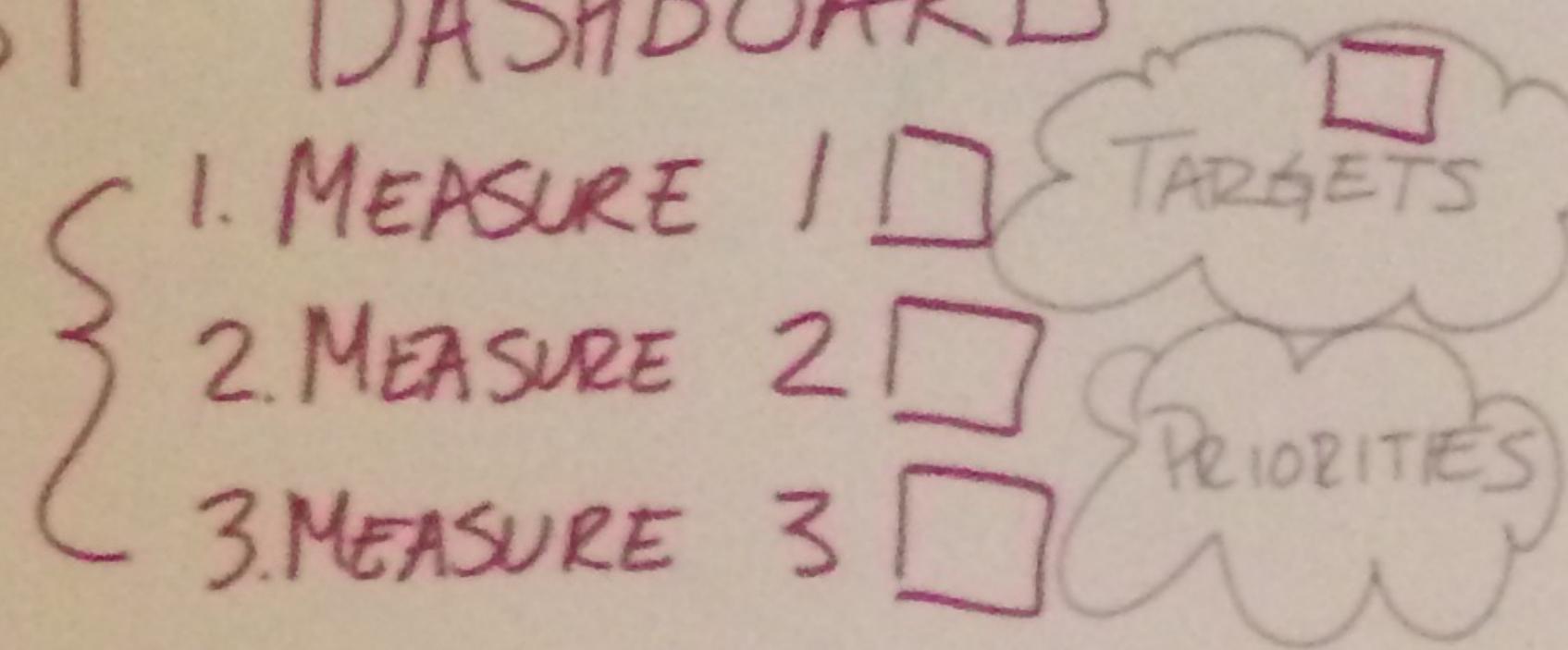
FRAMEWORK



GOALS

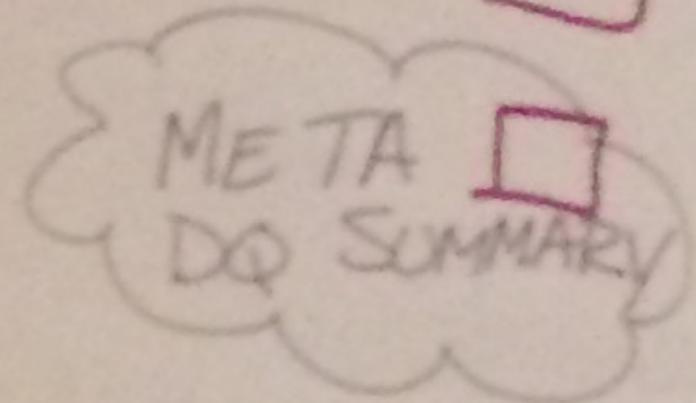
1. ANALYST DASHBOARD

3 DIFF
MEASURE
TYPES



2. ANALYST/RESEARCHER PI META DASHBOARD

1. SUMMARY OF MEASURES



DEM

L1

MS DATA

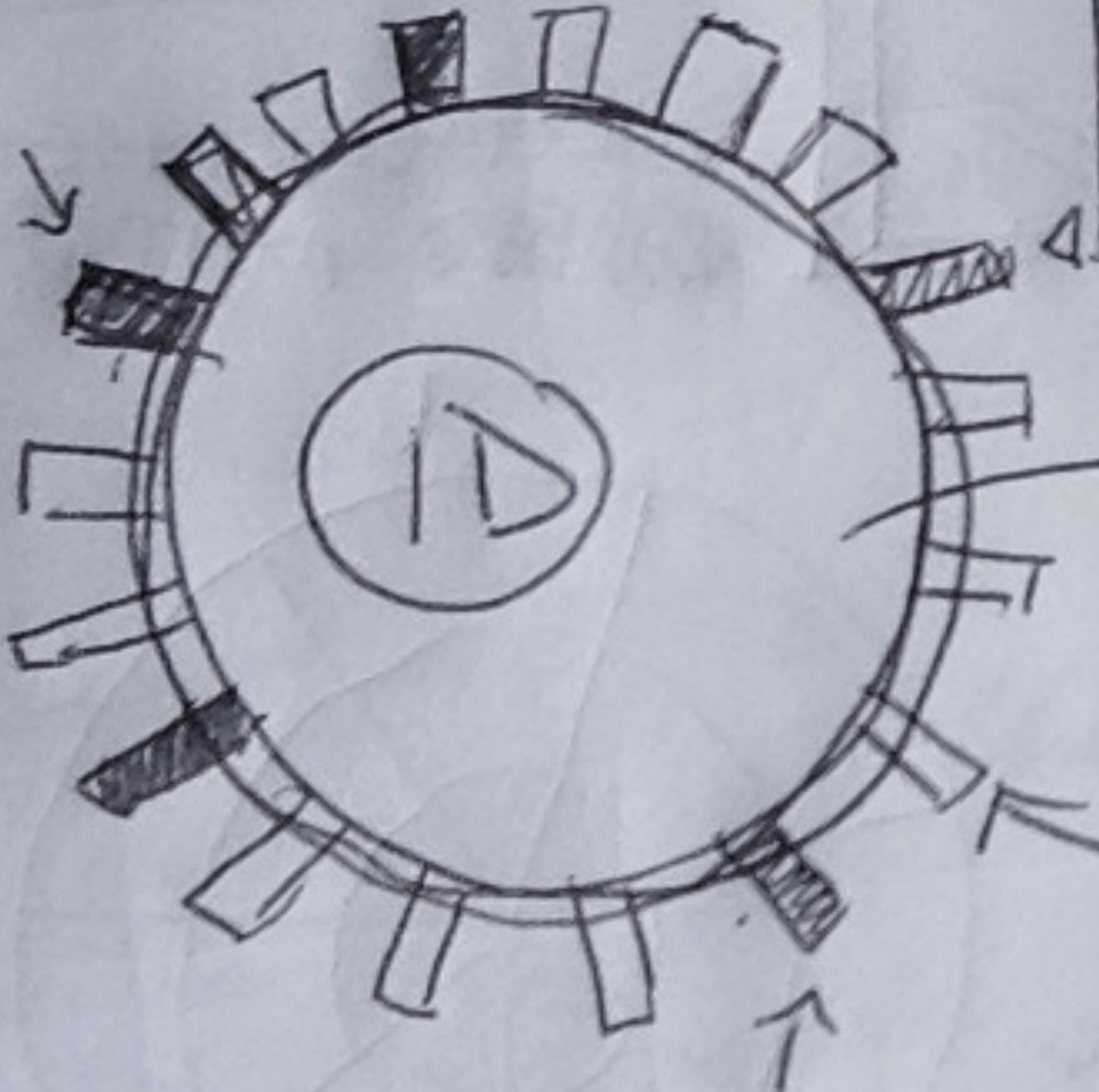
= Completeness
Content

100K
500K
MISSING
Content/
formats

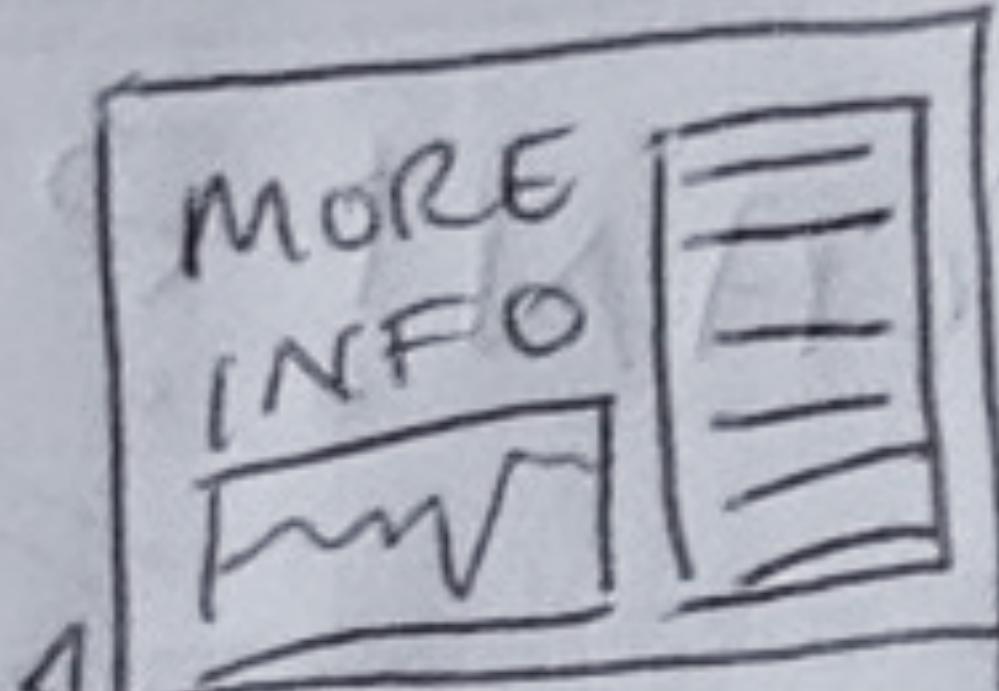
L2 = More complex
between vars
between tables

L3 = "higher level"
Tables of Freqs / X-freqs
Aggregate stats
variability

L4 - Further....
Freq of Preg Codes



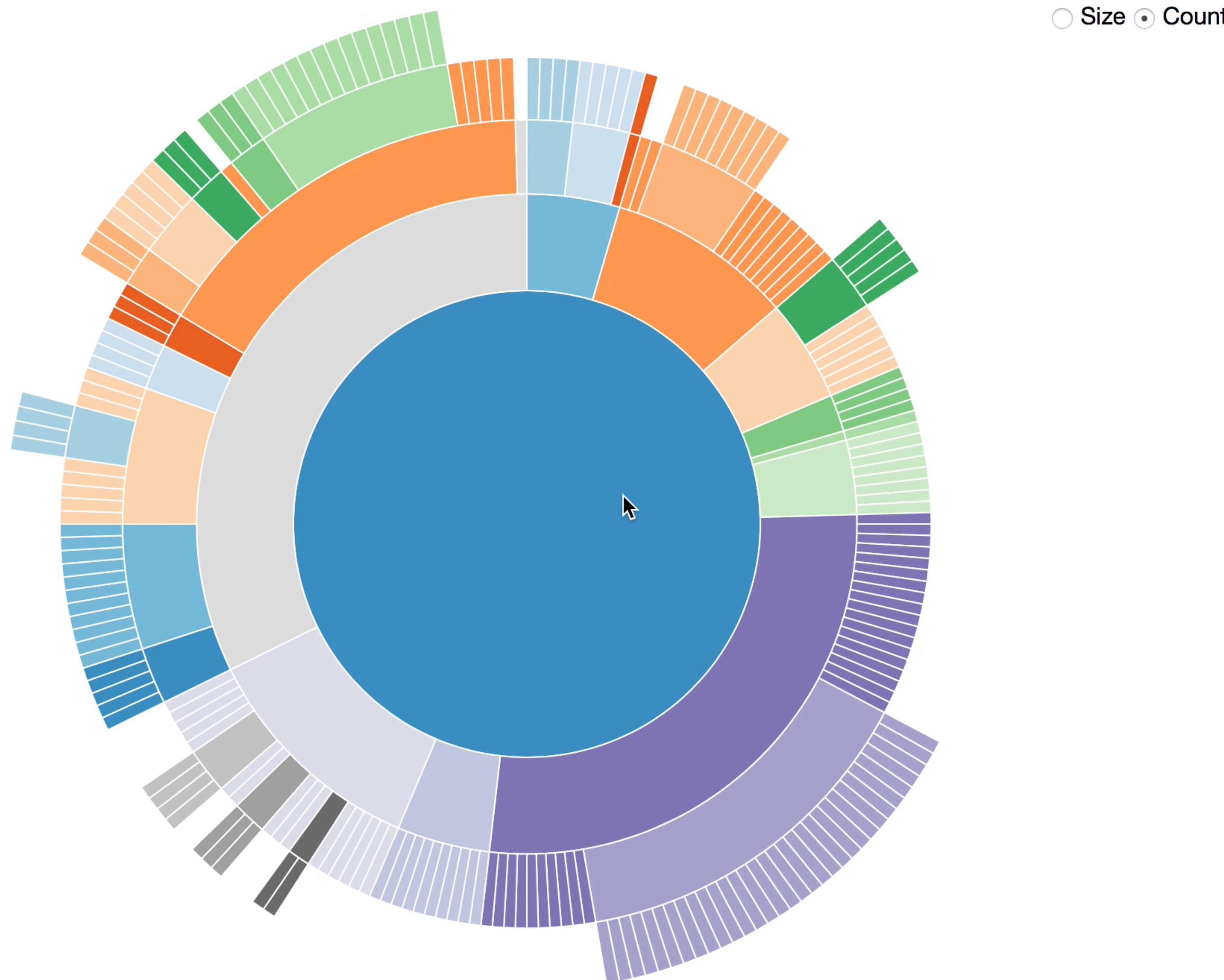
Color
coded



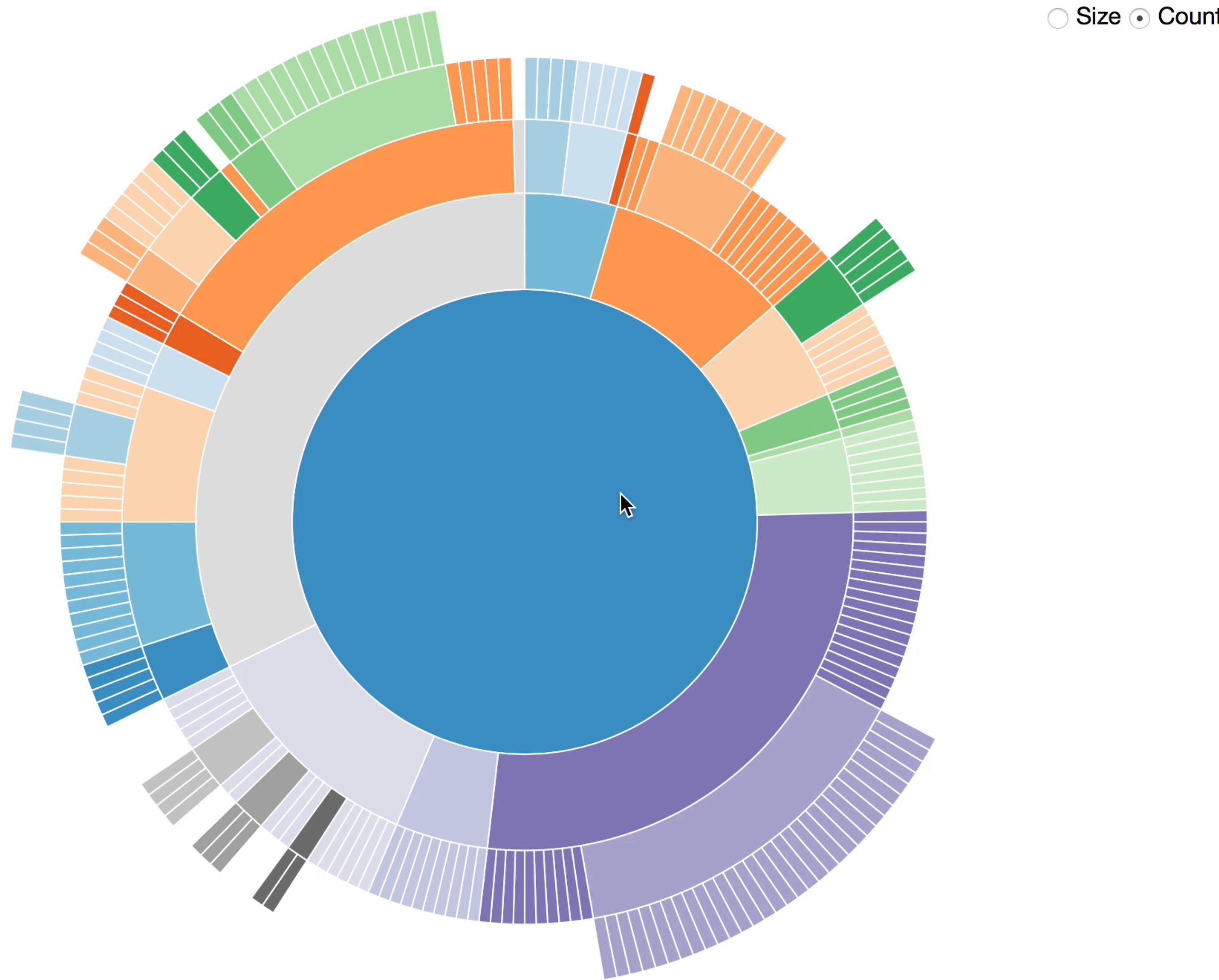
size corresponds to size of
data

each measure comes out
at 1 or 2 points of
circumferential

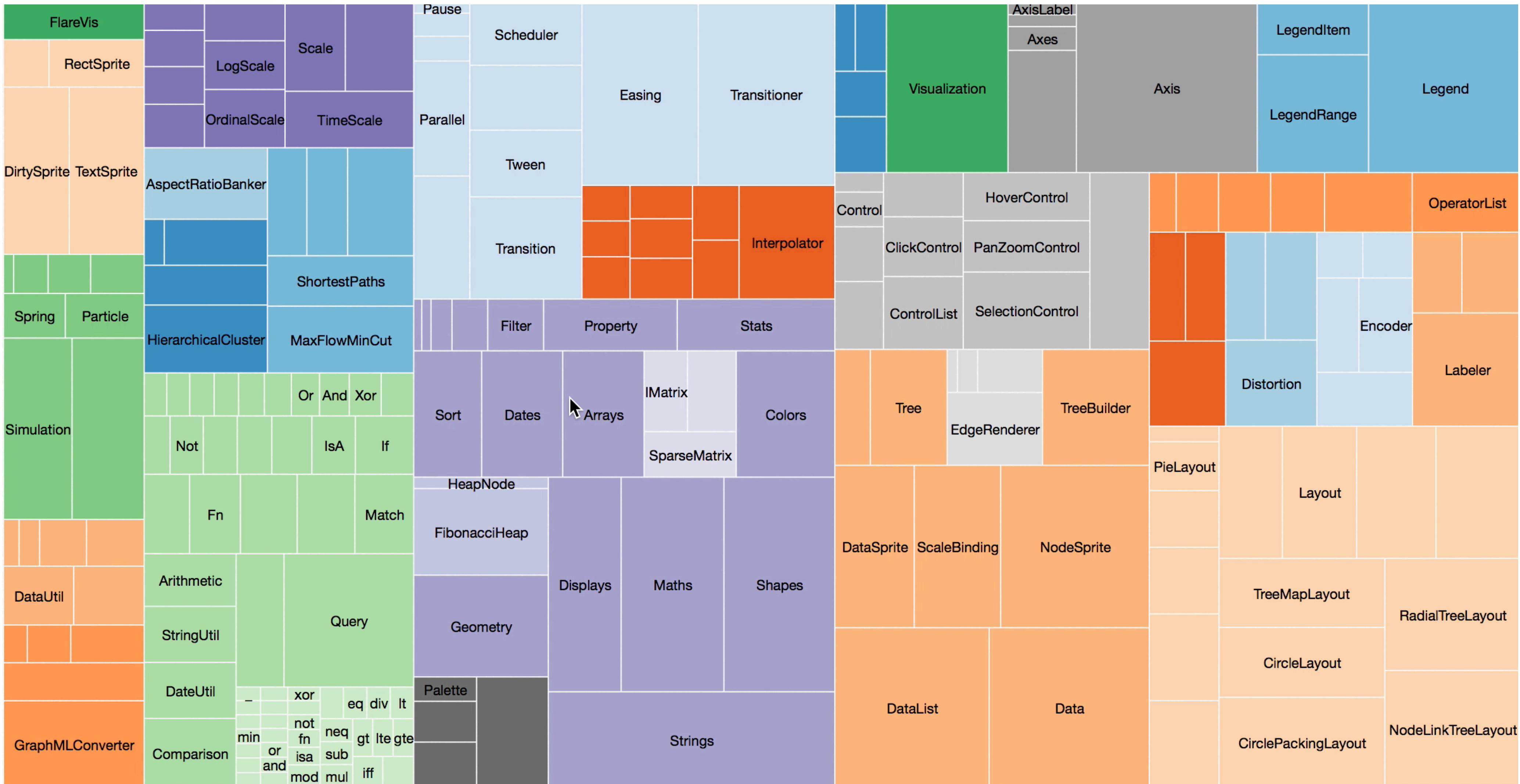
First, we wanted to make something like this.



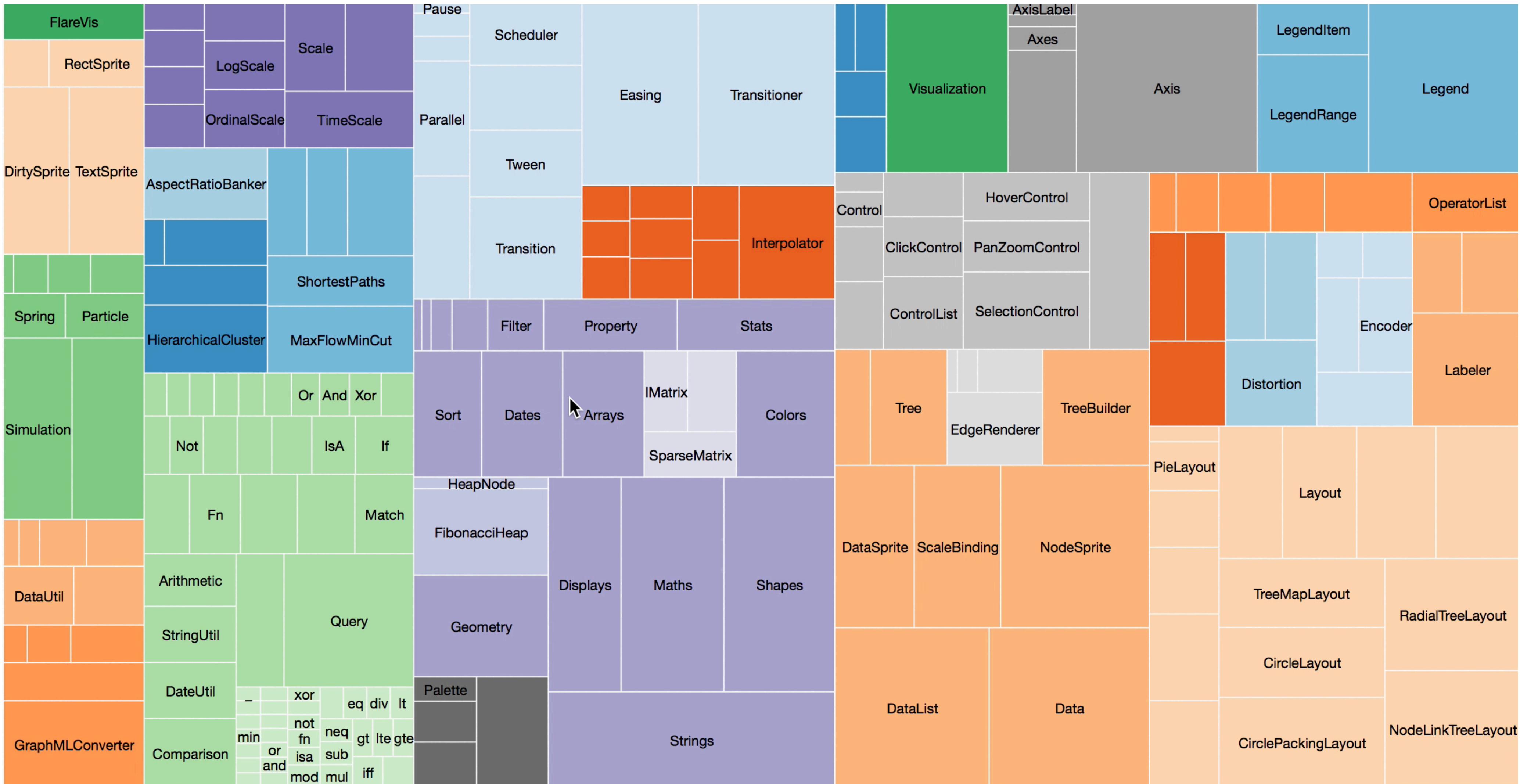
First, we wanted to make something like this.



Then, we wanted to make something like this.



Then, we wanted to make something like this.



- making use of
- ① CAN WE ACCESS THE MORE INDIVIDUALIZED PATIENT DATA?
- ② Generate our own statistics → differentiate between good + bad case scenarios
(makes more sense)

SANITY CHECKS !! (examples)

common discrepancies

- ① TOOL TO FIND ANOMALIES
- ② Ways TO GENERATE these anomalies
- ③ HOW TO DISPLAY THAT

341 340 342 343 344 345 346 347 348 349

350 351 352 353 354 355 356 357 358 359

360 361 362 363 364 365 366 367 368 369

370 371 372 373 374 375 376 377 378 379

380 381 382 383 384 385 386 387 388 389

390 391 392 393 394 395 396 397 398 399

400 401 402 403 404 405 406 407 408 409

410 411 412 413 414 415 416 417 418 419

420 421 422 423 424 425 426 427 428 429

430 431 432 433 434 435 436 437 438 439

440 441 442 443 444 445 446 447 448 449

450 451 452 453 454 455 456 457 458 459

460 461 462 463 464 465 466 467 468 469

470 471 472 473 474 475 476 477 478 479

480 481 482 483 484 485 486 487 488 489

490 491 492 493 494 495 496 497 498 499

500 501 502 503 504 505 506 507 508 509

510 511 512 513 514 515 516 517 518 519

520 521 522 523 524 525 526 527 528 529

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

580 581 582 583 584 585 586 587 588 589

590 591 592 593 594 595 596 597 598 599

600 601 602 603 604 605 606 607 608 609

610 611 612 613 614 615 616 617 618 619

620 621 622 623 624 625 626 627 628 629

630 631 632 633 634 635 636 637 638 639

640 641 642 643 644 645 646 647 648 649

650 651 652 653 654 655 656 657 658 659

660 661 662 663 664 665 666 667 668 669

670 671 672 673 674 675 676 677 678 679

680 681 682 683 684 685 686 687 688 689

690 691 692 693 694 695 696 697 698 699

700 701 702 703 704 705 706 707 708 709

710 711 712 713 714 715 716 717 718 719

720 721 722 723 724 725 726 727 728 729

730 731 732 733 734 735 736 737 738 739

740 741 742 743 744 745 746 747 748 749

750 751 752 753 754 755 756 757 758 759

760 761 762 763 764 765 766 767 768 769

770 771 772 773 774 775 776 777 778 779

780 781 782 783 784 785 786 787 788 789

790 791 792 793 794 795 796 797 798 799

800 801 802 803 804 805 806 807 808 809

810 811 812 813 814 815 816 817 818 819

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

920 921 922 923 924 925 926 927 928 929

930 931 932 933 934 935 936 937 938 939

940 941 942 943 944 945 946 947 948 949

950 951 952 953 954 955 956 957 958 959

960 961 962 963 964 965 966 967 968 969

970 971 972 973 974 975 976 977 978 979

980 981 982 983 984 985 986 987 988 989

990 991 992 993 994 995 996 997 998 999

1000 1001 1002 1003 1004 1005 1006 1007 1008 1009

1010 1011 1012 1013 1014 1015 1016 1017 1018 1019

1020 1021 1022 1023 1024 1025 1026 1027 1028 1029

1030 1031 1032 1033 1034 1035 1036 1037 1038 1039

1040 1041 1042 1043 1044 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 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 1277 1278 1279

1280 1281 1282 1283 1284 1285 1286 1287 1288 1289

1290 1291 1292 1293 1294 1295 1296 1297 1298 1299

1300 1301 1302 1303 1304 1305 1306 1307 1308 1309

1310 1311 1312 1313 1314 1315 1316 1317 1318 1319

1320 1321 1322 1323 1324 1325 1326 1327 1328 1329

1330 1331 1332 1333 1334 1335 1336 1337 1338 1339

1340 1341 1342 1343 1344 1345 1346 1347 1348 1349

1350 1351 1352 1353 1354 1355 1356 1357 1358 1359

1360 1361 1362 1363 1364 1365 1366 1367 1368 1369

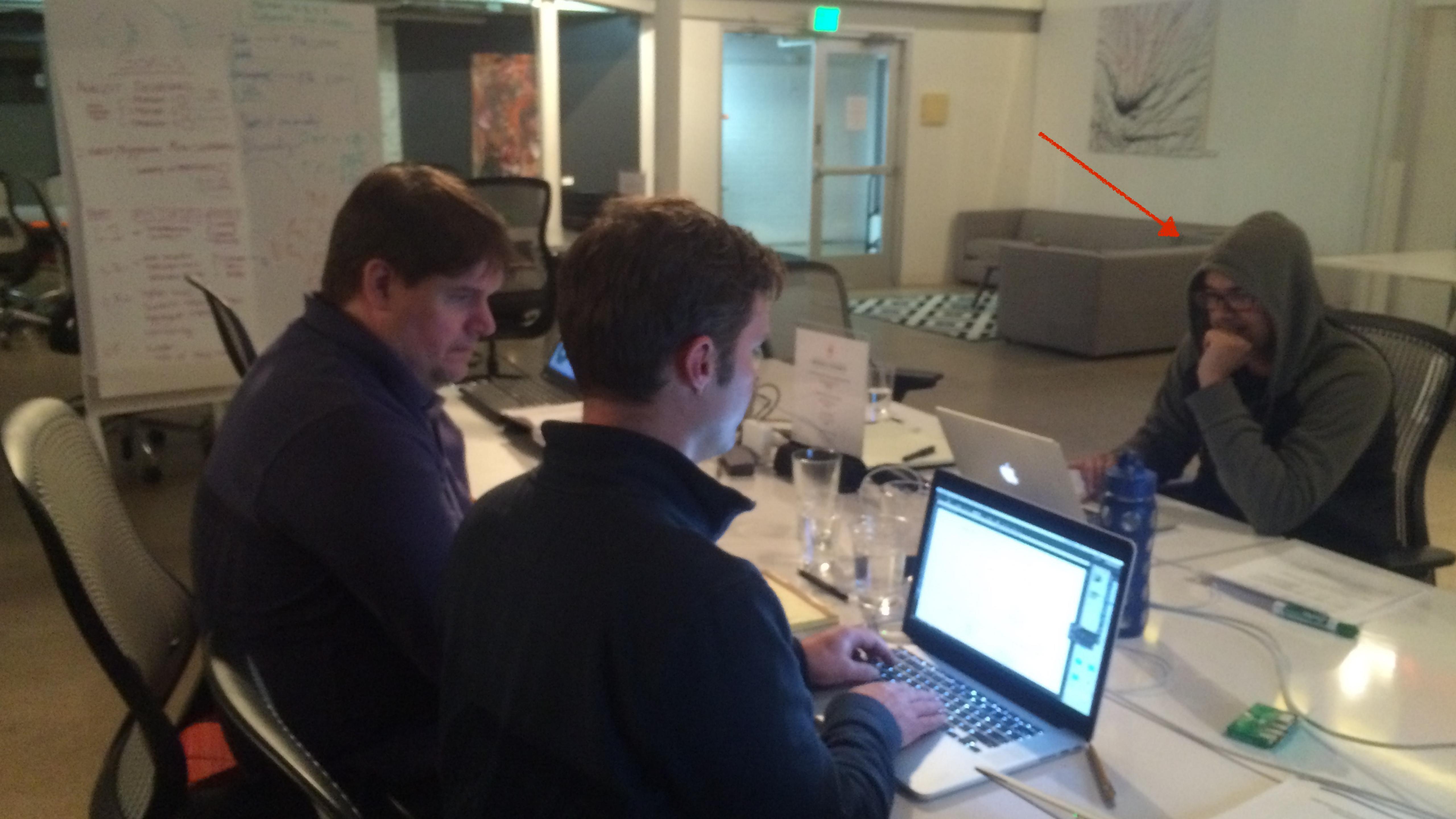
1370 1371 1372 1373 1374 1375 1376 1377 1378 1379

1380 1381 1382 1383 1384 1385 1386 1387 1388 1389

1390 1391 1392 1393 1394 1395 1396 1397 1398 1399

1400 1401 1402 1403 1404 1405 1406 1407 1408 1409

1410 1411 1412 1





868 observations of 3 variables

210	phis	1555	Px codes are invalid for >=.1% of cases.
211	phis	2264	Attending physician entered <=98% cases.
212	phis	1336	Financial data missing rate. (CTC Flag=No and >=1%)
213	phis	1867	Payer-Tertiary >= 1% unknown.
214	phis	1612	PDx code is invalid for >=.5% of cases.
215	phis	1997	Country code invalid value count
216	phis	2009	Reported to Chg based ECMO count difference >=3x median
217	phis	1222	Billing Number duplicate count
218	phis	2660	Discharge Hour = Medical Discharge Hour for 100% of cases
219	phis	1696	APGAR 10-minute score is missing for >=98% of neonates
220	phis	1324	Billed charges unmapped rate >\$50K or >.2% of billed chgs
221	phis	1275	CTC unspecified code charges-Supply
222	phis	1223	Encounter duplicate count (same BN & demographics)
223	phis	2155	PDx code = symptom for >2x PHIS median rate.
224	minisentinel	1	all_12_n_records_dupes.sas7bdat
225	minisentinel	2	cod_11_cod.sas7bdat
226	minisentinel	3	cod_12_patidmatch.sas7bdat
227	minisentinel	4	cod_13_causet.sas7bdat
228	minisentinel	5	cod_13_cod.sas7bdat
229	minisentinel	6	cod_13_codet.sas7bdat
230	minisentinel	7	cod_13_confidence.sas7bdat
231	minisentinel	8	cod_13_n_patid.sas7bdat
232	minisentinel	9	cod_13_source.sas7bdat
233	minisentinel	10	dem_12_patidmatch.sas7bdat
234	minisentinel	11	dem_13_ageyrsdist1.sas7bdat
235	minisentinel	12	dem_13_ageyrsdist2.sas7bdat



219,884 observations of 5 variables

	measure_id	set_id	measure_name	result_name	value
1	1	1	all_12_n_records_dupes.sas7bdat	name=COUNT	1.487070e+05
2	1	1	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	1.487070e+05
3	1	1	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
4	1	2	all_12_n_records_dupes.sas7bdat	name=COUNT	4.188355e+06
5	1	2	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	4.188355e+06
6	1	2	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
7	1	3	all_12_n_records_dupes.sas7bdat	name=COUNT	3.736901e+06
8	1	3	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	3.736901e+06
9	1	3	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
10	1	4	all_12_n_records_dupes.sas7bdat	name=COUNT	4.762000e+03
11	1	4	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	4.762000e+03
12	1	4	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
13	1	5	all_12_n_records_dupes.sas7bdat	name=COUNT	5.925453e+06
14	1	5	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	5.925453e+06
15	1	5	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
16	1	6	all_12_n_records_dupes.sas7bdat	name=COUNT	1.000000e+05
17	1	6	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	1.000000e+05
18	1	6	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
19	1	7	all_12_n_records_dupes.sas7bdat	name=COUNT	4.762000e+03
20	1	7	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	4.762000e+03
21	1	7	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	0.000000e+00
22	1	8	all_12_n_records_dupes.sas7bdat	name=COUNT	1.888325e+06
23	1	8	all_12_n_records_dupes.sas7bdat	name=COUNT DISTINCT	1.888011e+06
24	1	8	all_12_n_records_dupes.sas7bdat	name=PROBABILITY OF DUPLICATES	1.662849e-04

Displayed 1000 rows of 219,884 (218,884 omitted)

- can the measures be categorized?
- how to categorize the magnitude of the anomalies

HOW TO STRATIFY THE DATA?

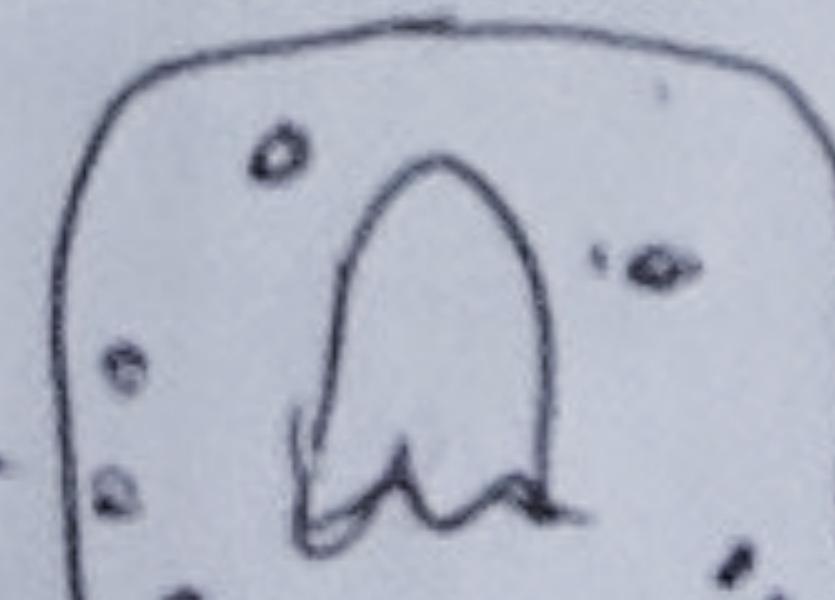
size vs count
↓
relative
children

WHAT IS THE SIZE OF THE
SQUARE REPRESENTING?
~~Categorization~~

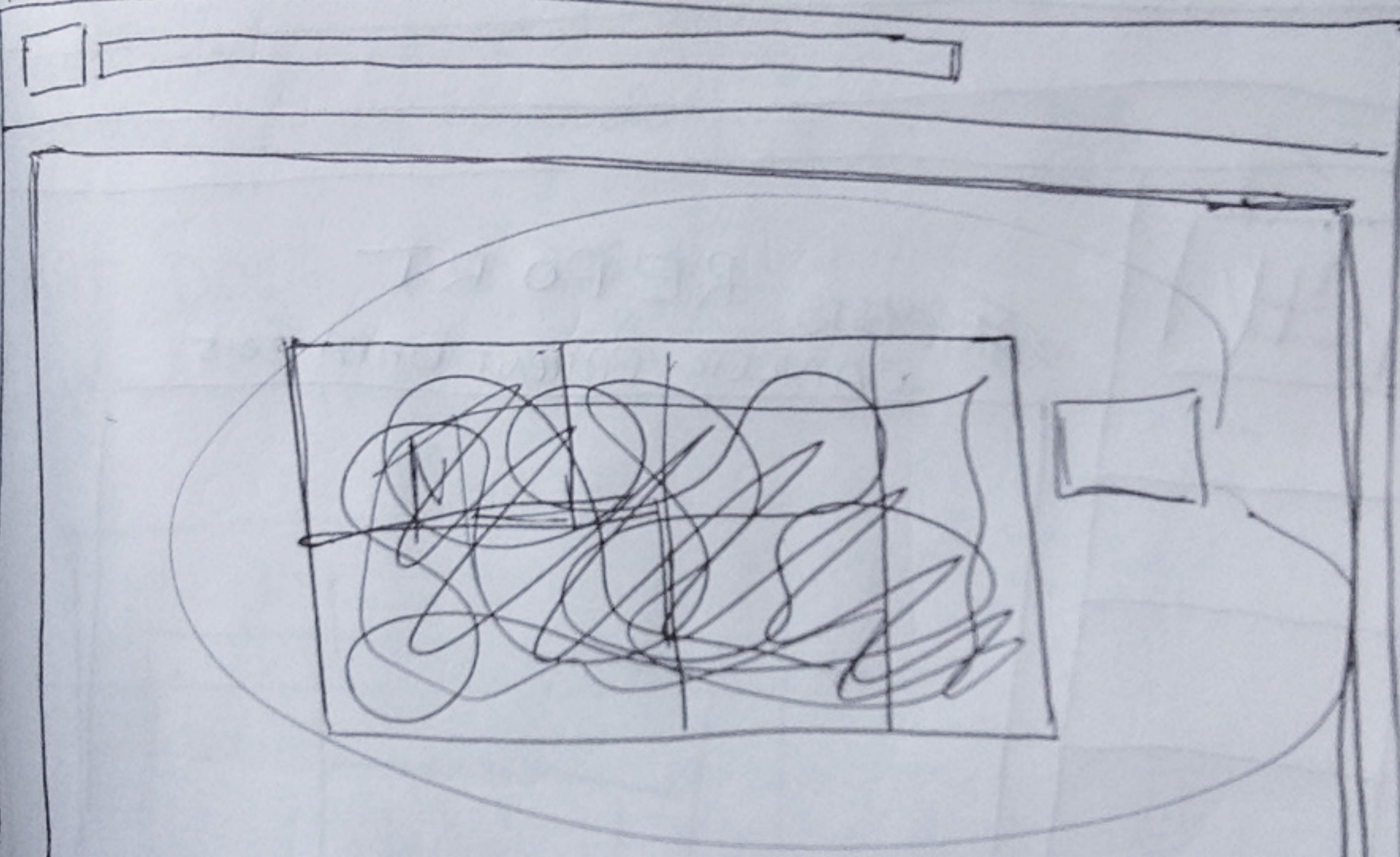
① WHAT QUADRANT?

②

Parse all ③ of
from



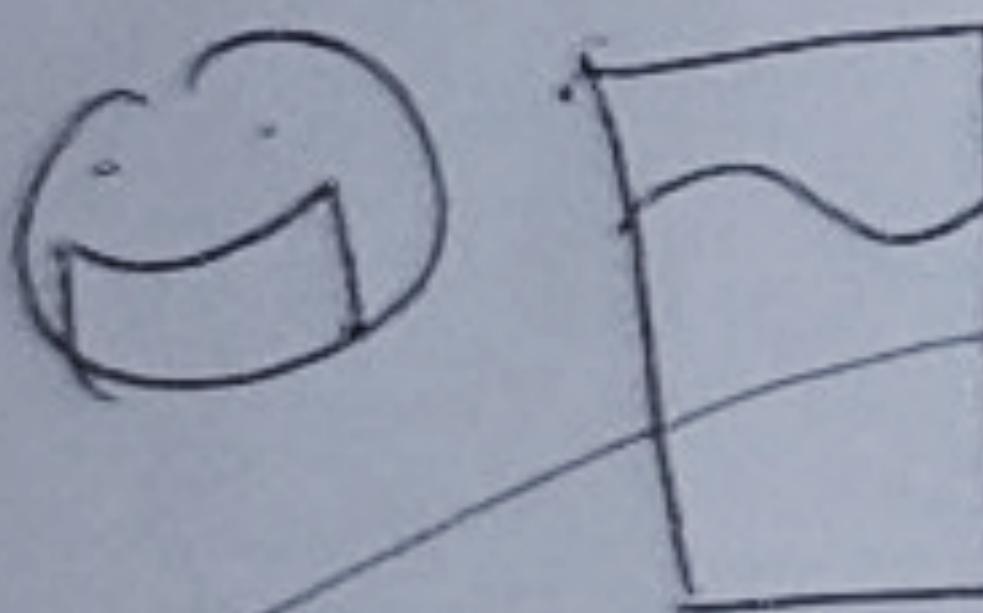
TOOL - ACCESSED THROUGH SOME SORT OF AC



84

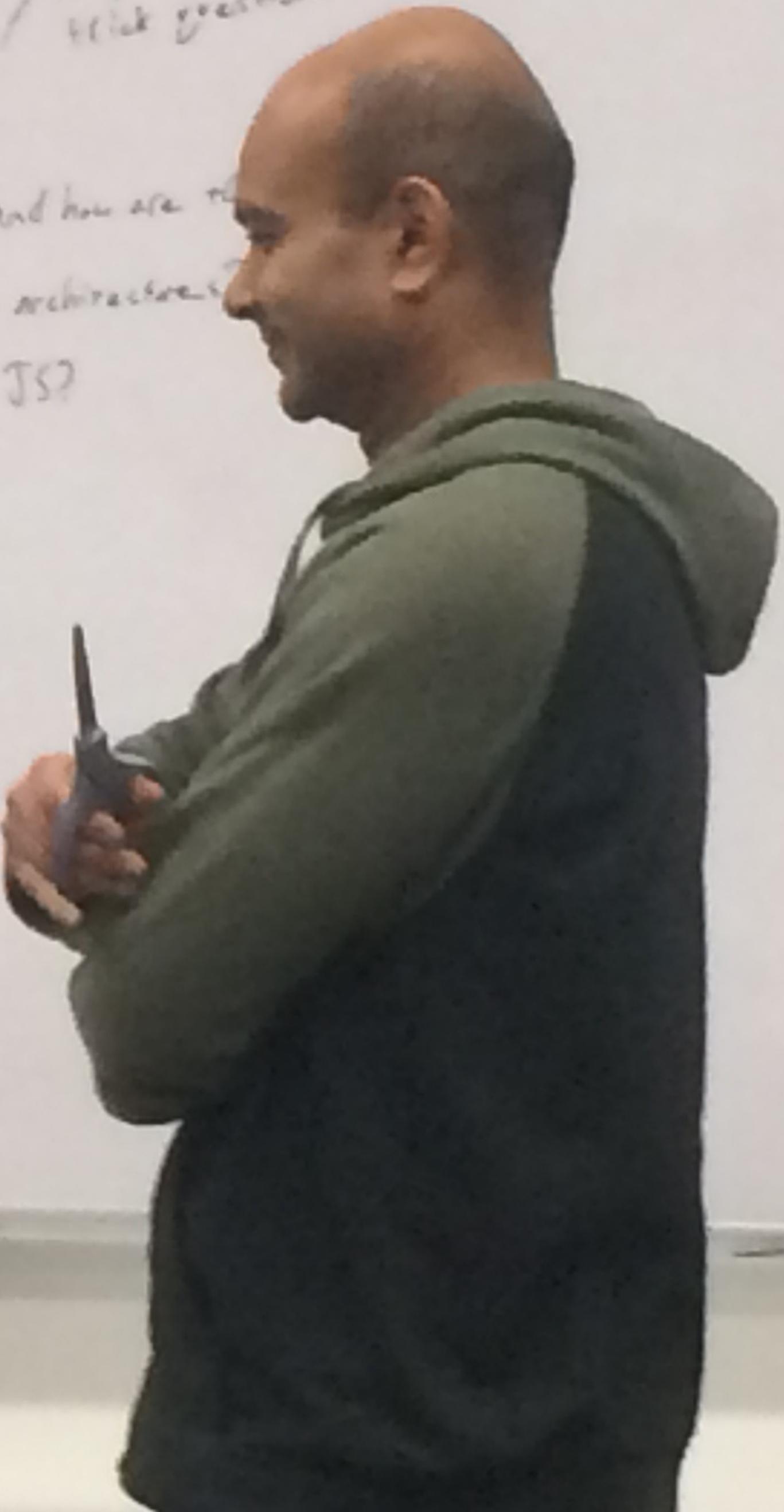
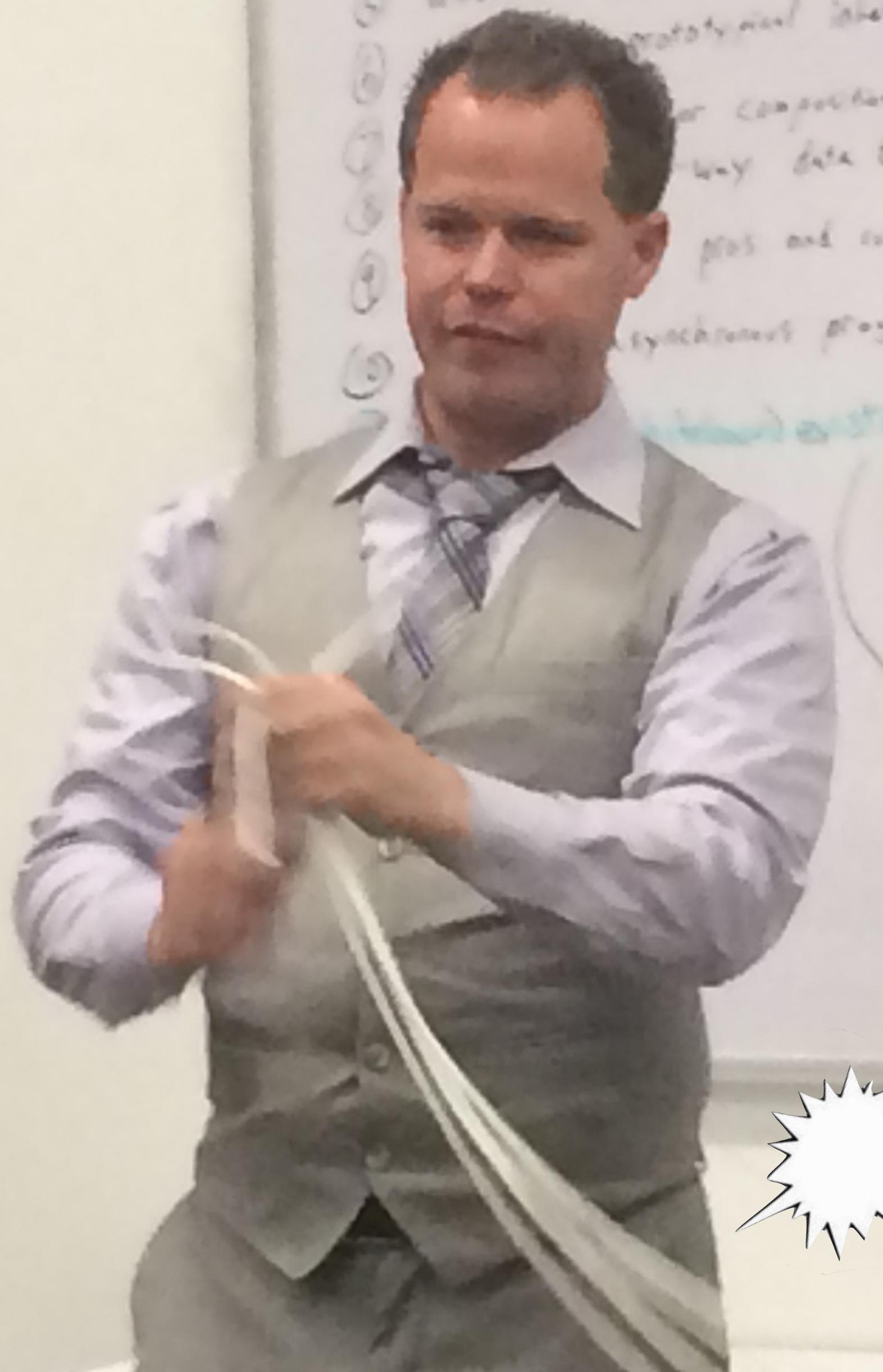
Level → Demographic → Anomaly

Beet!

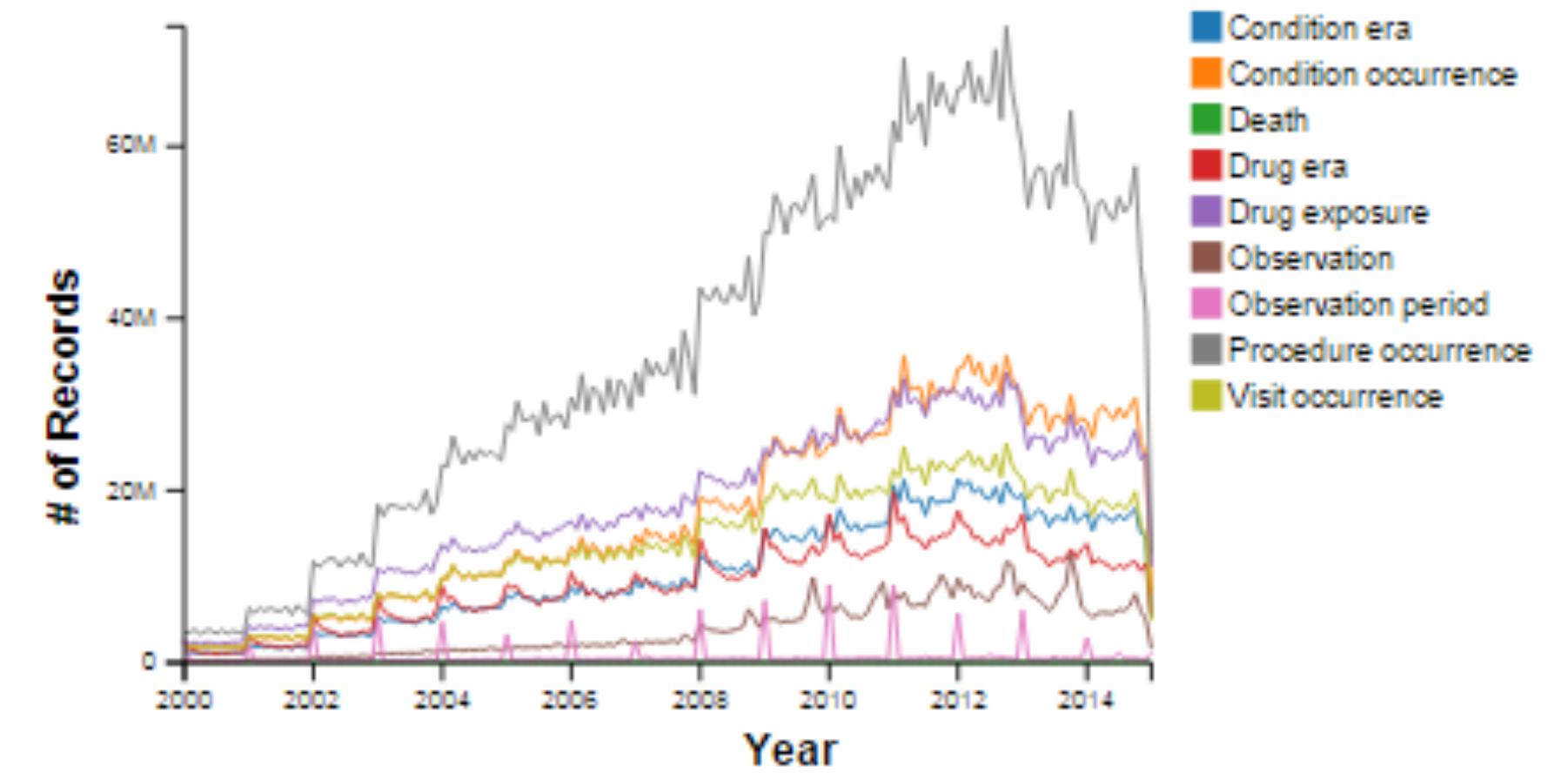


Despair!

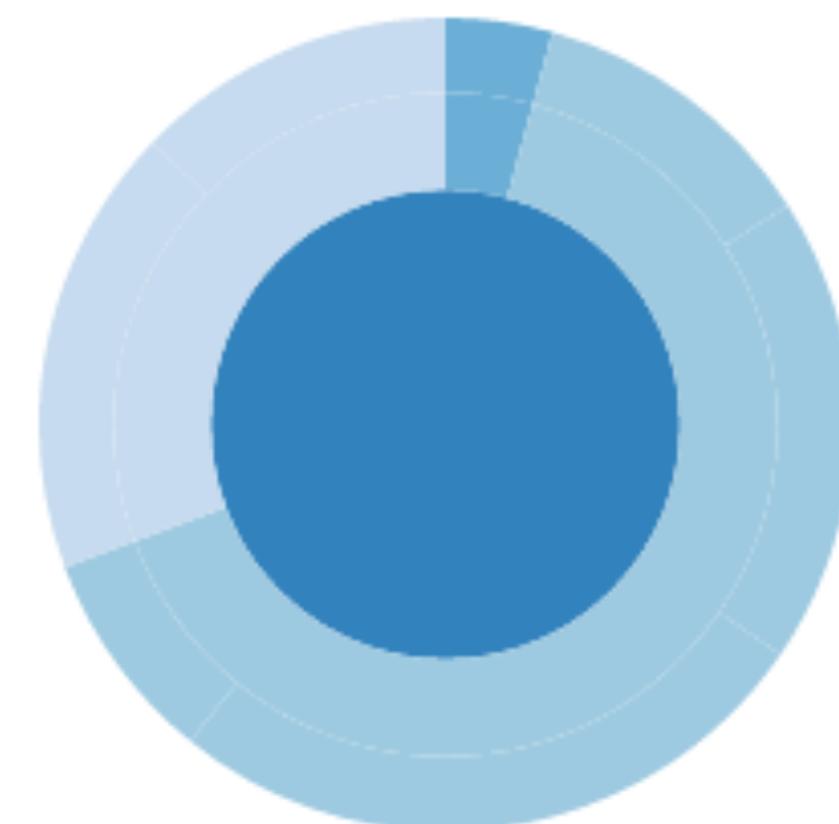
KNOW
THESE



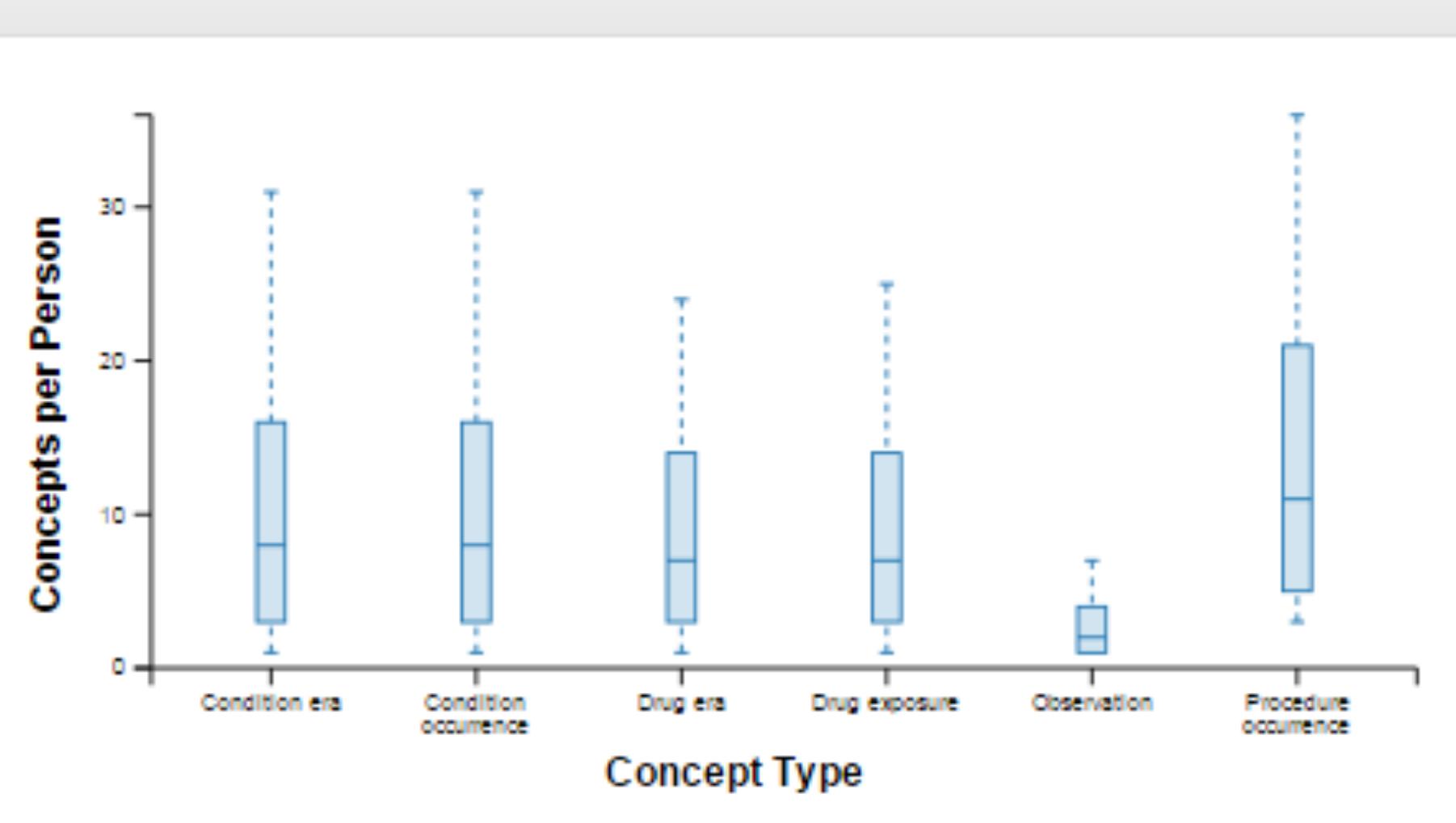
Multi-Series Line (Date Axis)



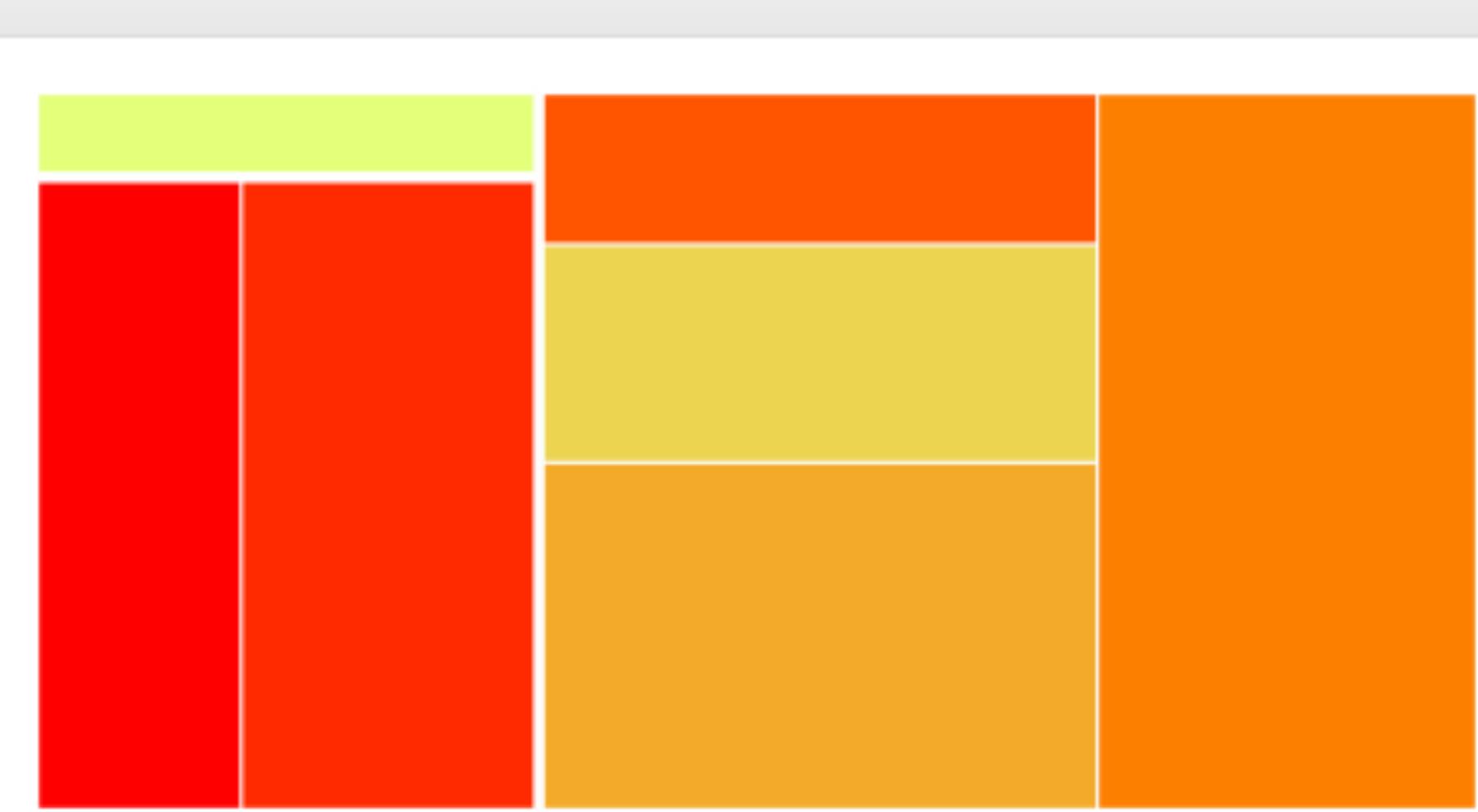
Sunburst



Boxplot



Treemap



And then, we made a prototype.

[HTTPS://INVIS.IO/GK4ULWS2E](https://invis.io/GK4ULWS2E)



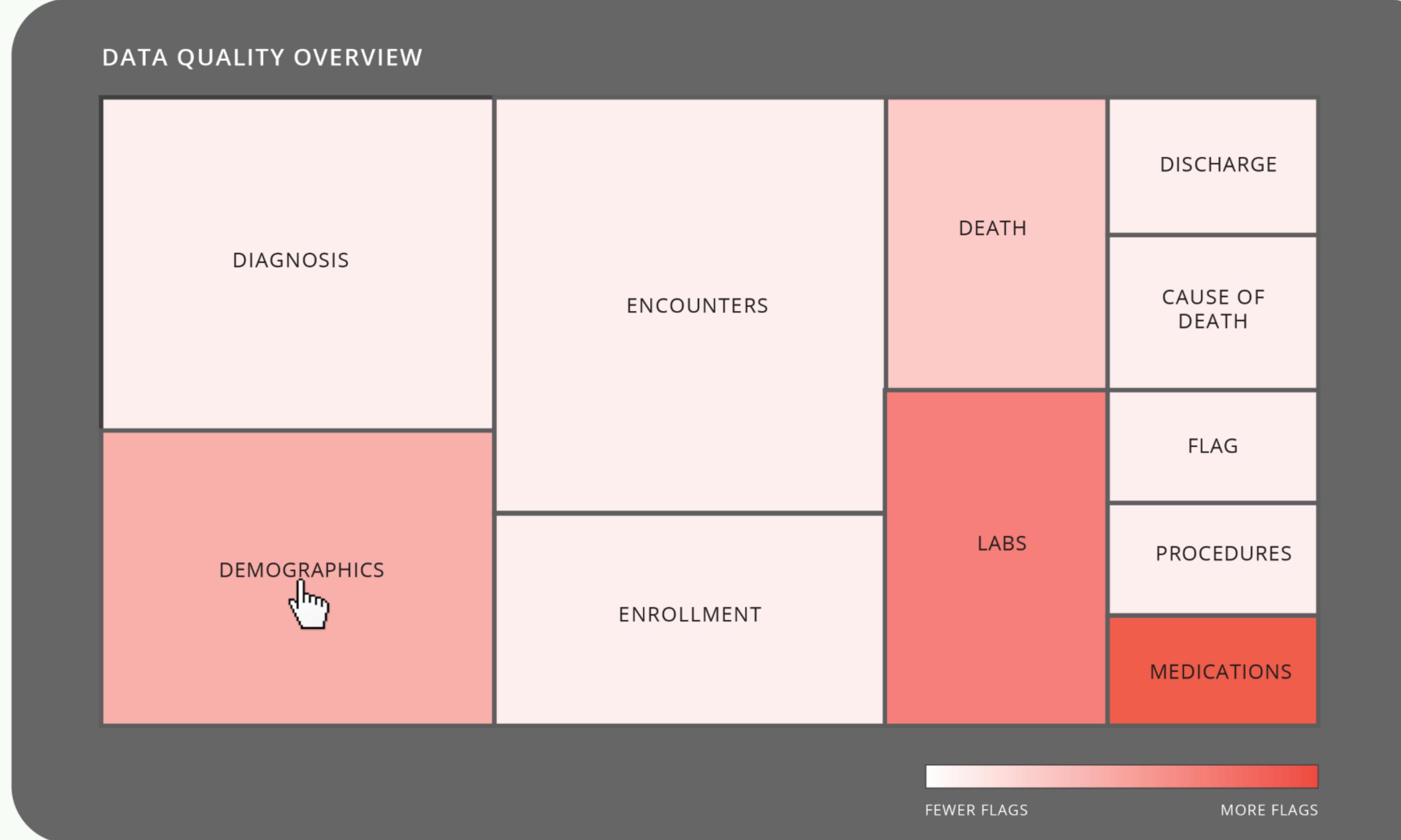
SUMMARY ANALYSIS

SUMMARY

COMPLETENESS

FIDELITY

PLAUSABILITY





DATA QUALITY MEASURE

SUMMARY

COMPLETENESS

FIDELITY

PLAUSABILITY

SUMMARY ANALYSIS

DEMOGRAPHIC DATA OVERVIEW

RACE

BIRTH YEAR

3 DIGIT ZIPCODE

SEX

ETHNICITY

FEWER FLAGS

MORE FLAGS

SUMMARY ANALYSIS

SUMMARY

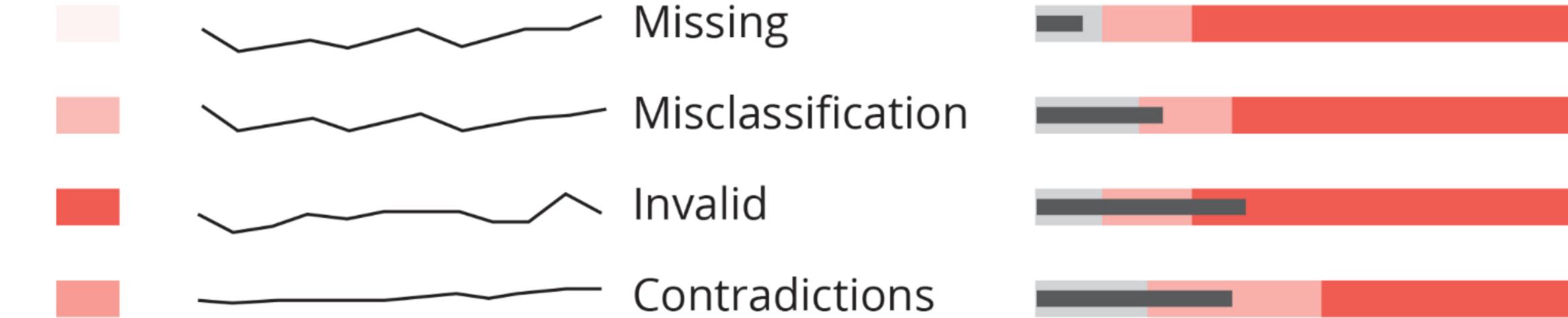
COMPLETENESS

FIDELITY

PLAUSABILITY

SEX ANALYSIS

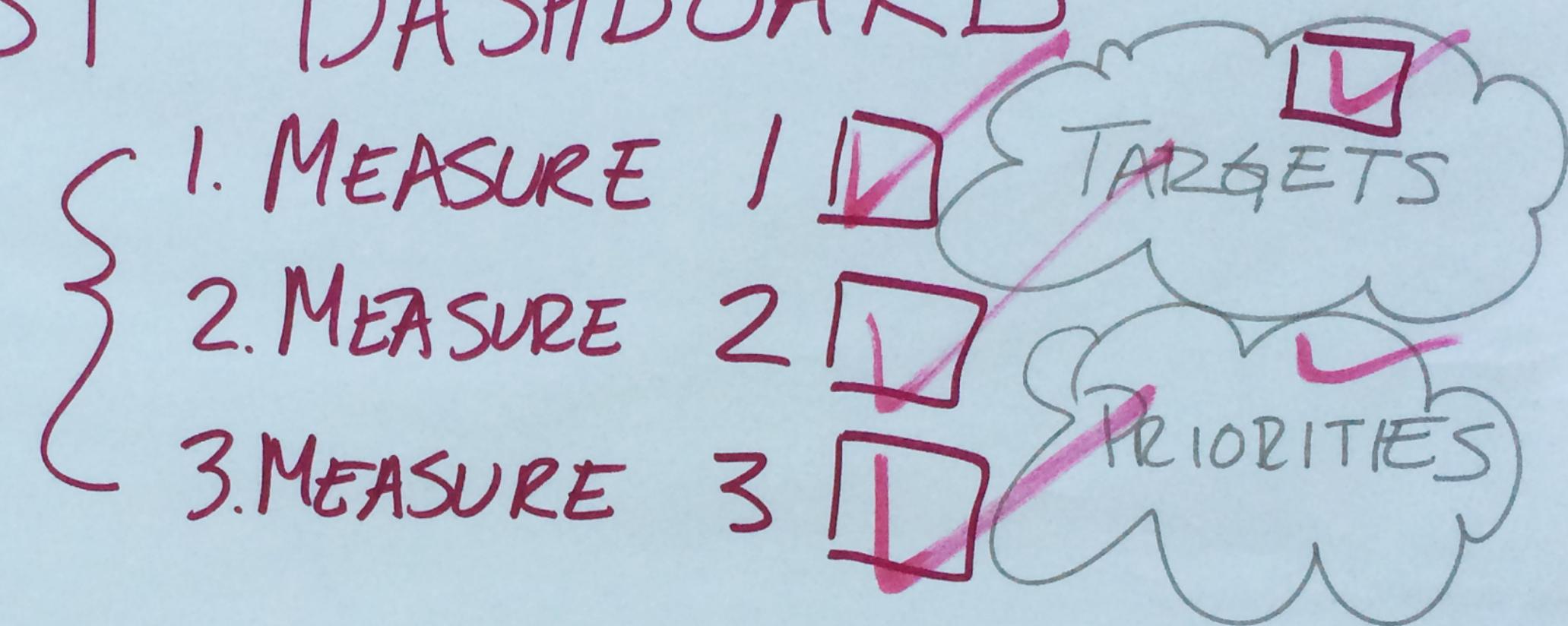
Past 12 Months Metric % of Records



GOALS

1. ANALYST DASHBOARD

3 DIFF
MEASURE
TYPES



2. ANALYST/RESEARCHER PI META DASHBOARD

1. SUMMARY OF MEASURES

→ FITNESS OF USE ←

