

Class 15: Pertussis mini-project

Youn Soo Na (PID: A1704731)

Background

Pertussis, a.k.a. whooping cough, is a highly infectious lung disease caused by the bacteria *B. Pertussis*.

The CDC tracks pertussis case numbers per year. Let's have a closer look at this [data](#):

We will use the `datapasta` R package to “scrape” this data into R.

Q1. With the help of the R “addin” package `datapasta` assign the CDC pertussis case number data to a data frame called `cdc` and use `ggplot` to make a plot of cases numbers over time.

```
library(ggplot2)
library(datapasta)

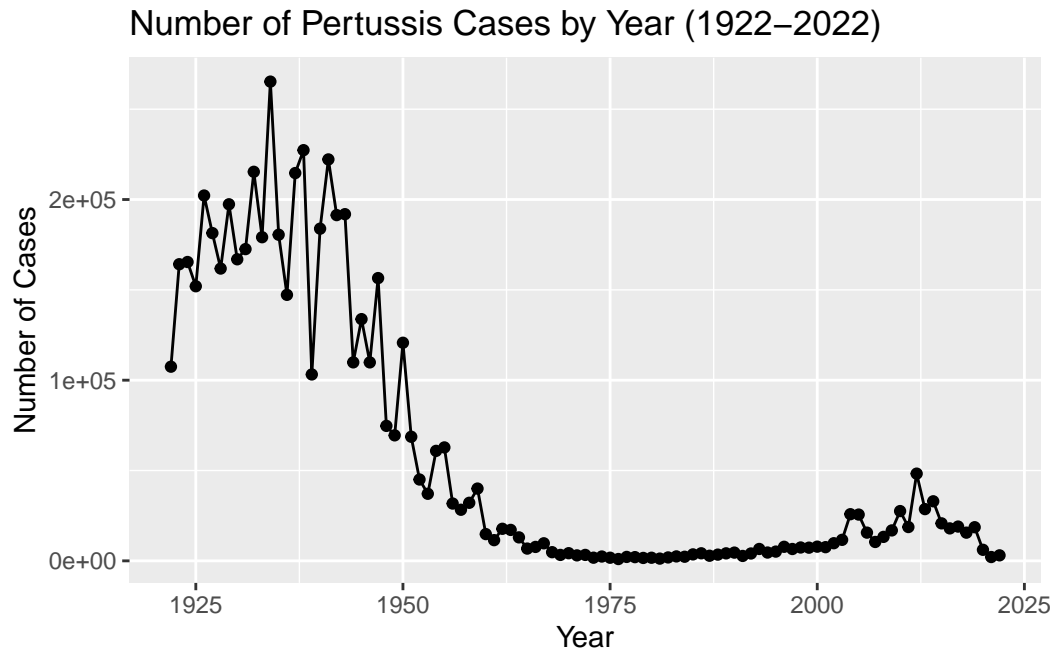
cdc <- data.frame(
  Year = c(1922L,
           1923L, 1924L, 1925L, 1926L, 1927L, 1928L,
           1929L, 1930L, 1931L, 1932L, 1933L, 1934L, 1935L,
           1936L, 1937L, 1938L, 1939L, 1940L, 1941L,
           1942L, 1943L, 1944L, 1945L, 1946L, 1947L, 1948L,
           1949L, 1950L, 1951L, 1952L, 1953L, 1954L,
           1955L, 1956L, 1957L, 1958L, 1959L, 1960L,
           1961L, 1962L, 1963L, 1964L, 1965L, 1966L, 1967L,
           1968L, 1969L, 1970L, 1971L, 1972L, 1973L,
           1974L, 1975L, 1976L, 1977L, 1978L, 1979L, 1980L,
           1981L, 1982L, 1983L, 1984L, 1985L, 1986L,
           1987L, 1988L, 1989L, 1990L, 1991L, 1992L, 1993L,
           1994L, 1995L, 1996L, 1997L, 1998L, 1999L,
           2000L, 2001L, 2002L, 2003L, 2004L, 2005L,
           2006L, 2007L, 2008L, 2009L, 2010L, 2011L, 2012L,
           2013L, 2014L, 2015L, 2016L, 2017L, 2018L,
```

```

2019L,2020L,2021L,2022L),
No..Reported.Pertussis.Cases = c(107473,
164191,165418,152003,202210,181411,
161799,197371,166914,172559,215343,179135,
265269,180518,147237,214652,227319,103188,
183866,222202,191383,191890,109873,
133792,109860,156517,74715,69479,120718,
68687,45030,37129,60886,62786,31732,28295,
32148,40005,14809,11468,17749,17135,
13005,6799,7717,9718,4810,3285,4249,
3036,3287,1759,2402,1738,1010,2177,2063,
1623,1730,1248,1895,2463,2276,3589,
4195,2823,3450,4157,4570,2719,4083,6586,
4617,5137,7796,6564,7405,7298,7867,
7580,9771,11647,25827,25616,15632,10454,
13278,16858,27550,18719,48277,28639,
32971,20762,17972,18975,15609,18617,6124,
2116,3044)
)

baseplot <- ggplot(cdc, aes(Year, No..Reported.Pertussis.Cases)) +
  geom_point() +
  geom_line() +
  labs(x="Year", y = "Number of Cases", title = "Number of Pertussis Cases by Year (1922-2022)")
baseplot

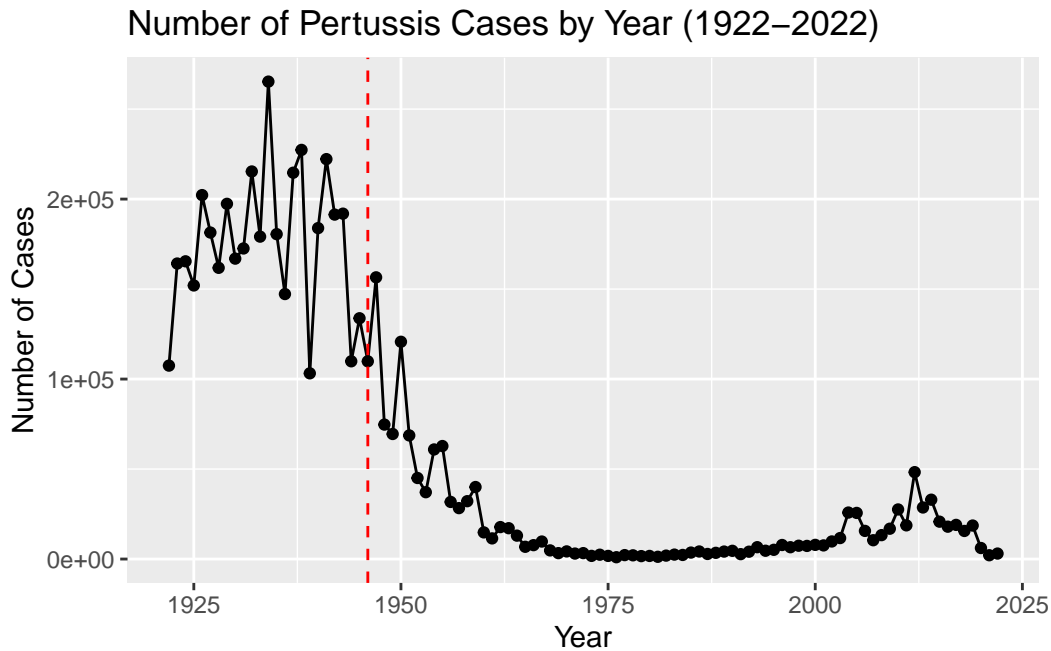
```



Add some landmark development as annotation to our plot. We include the first whole-cell (wP) vaccine roll-out in 1946.

```
baseplot +
  geom_vline(xintercept = 1946, linetype = "dashed", col = "red") +
  geom_text(aes(x = 1948, y = 1, label = "wP"), color = "red", vjust = -30)
```

Warning in geom_text(aes(x = 1948, y = 1, label = "wP"), color = "red", : All aesthetics have been used in the geom. Please consider using `annotate()` or provide this layer with data containing a single row.

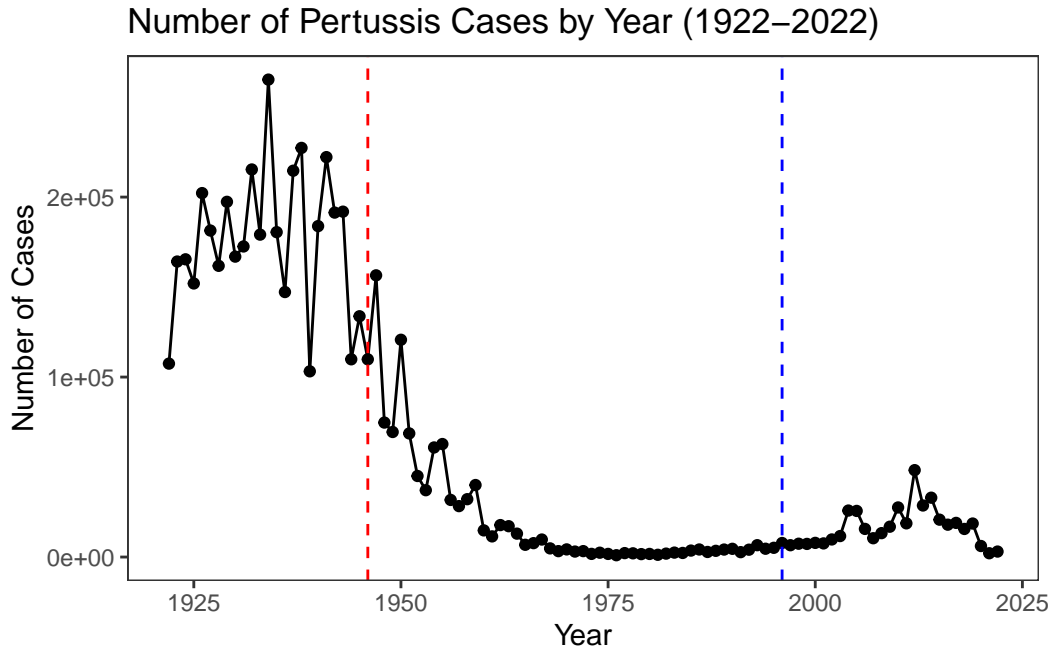


Let's add the switch to acellular vaccine (aP). > Q2. Using the ggplot `geom_vline()` function add lines to your previous plot for the 1946 introduction of the wP vaccine and the 1996 switch to aP vaccine (see example in the hint below). What do you notice?

```
baseplot +
  geom_vline(xintercept = 1946, linetype = "dashed", col = "red") +
  geom_vline(xintercept = 1996, linetype = "dashed", col = "blue") +
  geom_text(aes(x = 1998, y = 1, label = "aP"), color = "blue", vjust = -30) +
  geom_text(aes(x = 1948, y = 1, label = "wP"), color = "red", vjust = -30) +
  theme_test()
```

Warning in `geom_text(aes(x = 1998, y = 1, label = "aP"), color = "blue",` : All aesthetics have been mapped to the same value. Please consider using ``annotate()`` or provide this layer with data containing a single row.

Warning in `geom_text(aes(x = 1948, y = 1, label = "wP"), color = "red",` : All aesthetics have been mapped to the same value. Please consider using ``annotate()`` or provide this layer with data containing a single row.



Q3. Describe what happened after the introduction of the aP vaccine? Do you have a possible explanation for the observed trend?

We went from ~200,000 cases pre-wP vaccine to ~1,000 cases in 1976.

The US switched to the aP vaccine in 1995. We start to see a big increase in 2004 to ~26,000 cases.

There is a ~10 year lag from aP roll-out to increasing case numbers. This holds true of other countries like Japan, UK, etc. It seems that something about the aP vaccine that has a immune response that does not last as long as the wP vaccine.

Key Question: Why does the aP vaccine induced immunity wane faster than that of the wP vaccine?

CMI-PB

The CMI-PB (Computation Models of Immunity Pertussis Boost) makes available lost of data about the immune response to Pertussis booster vaccination.

Critically, it tracks wP and aP individuals over time to see how their immune response changes.

CMI-PB make all their data freely available via JSON format tables from their databas.

Let's read the first one of these tables:

```
library(jsonlite)

subject <- read_json("https://www.cmi-pb.org/api/v5/subject", simplifyVector = T)

head(subject)
```

	subject_id	infancy_vac	biological_sex	ethnicity	race
1	1	wP	Female	Not Hispanic or Latino	White
2	2	wP	Female	Not Hispanic or Latino	White
3	3	wP	Female	Unknown	White
4	4	wP	Male	Not Hispanic or Latino	Asian
5	5	wP	Male	Not Hispanic or Latino	Asian
6	6	wP	Female	Not Hispanic or Latino	White

	year_of_birth	date_of_boost	dataset
1	1986-01-01	2016-09-12	2020_dataset
2	1968-01-01	2019-01-28	2020_dataset
3	1983-01-01	2016-10-10	2020_dataset
4	1988-01-01	2016-08-29	2020_dataset
5	1991-01-01	2016-08-29	2020_dataset
6	1988-01-01	2016-10-10	2020_dataset

Q1. How many subjects are there in this dataset?

```
nrow(subject)
```

```
[1] 172
```

Q4. How many aP and wP infancy vaccinated subjects are in the dataset?

```
table(subject$infancy_vac)
```

```
aP wP
87 85
```

Q5. How many Male and Female subjects/patients are in the dataset?

```
table(subject$biological_sex)
```

Female	Male
112	60

Q6. What is the breakdown of race and biological sex (e.g. number of Asian females, White males etc...)?

```
table(subject$race, subject$biological_sex)
```

	Female	Male
American Indian/Alaska Native	0	1
Asian	32	12
Black or African American	2	3
More Than One Race	15	4
Native Hawaiian or Other Pacific Islander	1	1
Unknown or Not Reported	14	7
White	48	32

Q. Does this do a good job of representing the US populus?

No way

Let's get more data from CMI-PB, this time about the specimens collected.

```
specimen <- read_json("https://cmi-pb.org/api/v5/specimen", simplifyVector = T)
```

Now we can join (merge) these two tables `subject` and `specimen` to make one new `meta` table with the combined data.

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

`filter`, `lag`

The following objects are masked from 'package:base':

`intersect`, `setdiff`, `setequal`, `union`

```
meta <- inner_join(subject, specimen)
```

Joining with `by = join_by(subject_id)`

```
head(meta)
```

	subject_id	infancy_vac	biological_sex	ethnicity	race
1	1	wP	Female	Not Hispanic or Latino	White
2	1	wP	Female	Not Hispanic or Latino	White
3	1	wP	Female	Not Hispanic or Latino	White
4	1	wP	Female	Not Hispanic or Latino	White
5	1	wP	Female	Not Hispanic or Latino	White
6	1	wP	Female	Not Hispanic or Latino	White

	year_of_birth	date_of_boost	dataset	specimen_id
1	1986-01-01	2016-09-12	2020_dataset	1
2	1986-01-01	2016-09-12	2020_dataset	2
3	1986-01-01	2016-09-12	2020_dataset	3
4	1986-01-01	2016-09-12	2020_dataset	4
5	1986-01-01	2016-09-12	2020_dataset	5
6	1986-01-01	2016-09-12	2020_dataset	6

	actual_day_relative_to_boost	planned_day_relative_to_boost	specimen_type
1	-3	0	Blood
2	1	1	Blood
3	3	3	Blood
4	7	7	Blood
5	11	14	Blood
6	32	30	Blood

	visit
1	1
2	2
3	3
4	4
5	5
6	6

Now read an “experiment data” table from CMI-PB

```
abdata <- read_json("https://www.cmi-pb.org/api/v5/plasma_ab_titer", simplifyVector = T)
head(abdata)
```


	specimen_id	isotype	is_antigen_specific	antigen	MFI	MFI_normalised
1	1	IgE	FALSE	Total	1110.21154	2.493425
2	1	IgE	FALSE	Total	2708.91616	2.493425
3	1	IgG	TRUE	PT	68.56614	3.736992
4	1	IgG	TRUE	PRN	332.12718	2.602350
5	1	IgG	TRUE	FHA	1887.12263	34.050956
6	1	IgE	TRUE	ACT	0.10000	1.000000

	unit	lower_limit_of_detection
1	UG/ML	2.096133
2	IU/ML	29.170000
3	IU/ML	0.530000
4	IU/ML	6.205949
5	IU/ML	4.679535
6	IU/ML	2.816431

Q. What are the different \$dataset values in abdata and what do you notice about the number of rows for the most “recent” dataset?

```
str(abdata)
```

```
'data.frame': 52576 obs. of 8 variables:
 $ specimen_id      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ isotype          : chr  "IgE" "IgE" "IgG" "IgG" ...
 $ is_antigen_specific : logi FALSE FALSE TRUE TRUE TRUE TRUE ...
 $ antigen          : chr  "Total" "Total" "PT" "PRN" ...
 $ MFI              : num  1110.2 2708.9 68.6 332.1 1887.1 ...
 $ MFI_normalised    : num  2.49 2.49 3.74 2.6 34.05 ...
 $ unit             : chr  "UG/ML" "IU/ML" "IU/ML" "IU/ML" ...
 $ lower_limit_of_detection: num  2.1 29.17 0.53 6.21 4.68 ...
```

```
table(abdata$specimen_id)
```

1	2	3	4	5	6	7	19	20	21	22	23	24	25	27	28
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
29	30	31	32	33	37	38	39	40	41	42	43	45	46	47	48
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
49	50	51	55	56	57	58	59	60	61	70	71	72	73	74	75
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
76	77	78	79	80	81	82	87	88	89	90	91	92	93	96	97
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80

98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
114	115	116	117	118	119	120	121	122	123	124	125	126	127	131	132
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
133	134	135	136	137	138	139	140	141	142	143	144	146	147	148	149
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
182	183	184	185	186	187	191	192	193	194	195	196	197	201	202	203
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
204	205	206	207	208	209	210	211	212	213	214	216	217	218	219	220
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
221	222	223	224	225	226	227	228	229	232	233	234	235	236	237	238
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
257	258	259	260	261	266	267	268	269	270	271	272	274	275	276	277
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
278	279	280	281	282	283	284	285	286	288	289	290	291	292	293	294
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
295	296	297	298	299	300	301	302	303	304	305	306	310	311	312	313
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
330	332	333	334	335	336	337	338	342	343	344	345	346	347	348	349
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
366	369	370	371	372	373	374	375	376	377	378	379	380	381	382	385
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
402	403	405	406	407	408	409	410	411	412	413	414	415	416	417	418
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
419	420	421	422	423	424	425	427	428	429	430	431	432	433	434	435
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
436	437	438	439	440	441	442	443	444	445	446	447	450	451	452	453
80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80
454	455	456	458	459	460	461	462	463	464	468	469	470	471	472	473
80	80	80	80	80	80	80	80	80	80	35	35	35	35	35	35
474	475	476	477	478	479	480	481	483	484	485	486	487	488	489	490

35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
491	492	493	494	495	496	498	499	500	501	502	503	504	506	507	508
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
509	510	511	512	513	514	515	516	517	518	519	521	522	523	524	525
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
526	527	529	530	531	532	533	534	535	537	538	539	540	541	542	543
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
546	547	548	549	550	551	552	554	555	556	557	558	559	560	562	563
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
564	565	566	567	568	569	570	571	572	573	574	575	577	578	579	580
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
581	582	583	585	586	587	588	589	590	591	593	594	595	596	597	598
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
599	601	602	603	604	605	606	607	608	609	610	611	612	613	614	616
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
617	618	619	620	621	622	623	624	625	626	627	628	629	636	637	638
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
655	656	657	658	659	660	661	662	663	674	675	676	677	678	679	680
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
809	810	811	812	813	814	815	816	817	818	820	821	822	823	824	825
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35

858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
938	939	940	942	943	955	956	957	962	965	966	973	976	977	984	986
35	21	35	35	35	35	35	35	35	35	35	35	35	35	35	35
987	994	996	997	1004	1006	1007	1014	1017	1018	1025	1028	1029	1036	1037	1038
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1045	1048	1049	1056	1057	1058	1064	1066	1067	1074	1077	1078	1085	1088	1089	1096
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1097	1098	1105	1107	1108	1115	1118	1119	1126	1129	1130	1137	1140	1141	1148	1151
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1152	1159	1162	1163	1170	1173	1174	1181	1183	1184	1191	1194	1195	1202	1205	1206
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1213	1216	1217	1224	1226	1227	1234	1237	1238	1245	1247	1248	1255	1257	1258	1265
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1268	1269	1276	1278	1279	1286	1289	1290	1297	1299	1300	1307	1310	1311	1318	1321
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1322	1329	1331	1332	1339	1342	1343	1350	1351	1352	1359	1362	1363	1370	1373	1374
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1381	1384	1385	1392	1395	1396	1403	1406	1407	1414	1416	1417	1424	1426	1427	1434
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1436	1437	1444	1446	1447	1454	1456	1457	1464	1466	1467	1474	1476	1477	1484	1486
35	35	35	35	35	35	35	35	35	35	35	35	35	35	35	35
1487	1494	1496	1497												
35	35	35	35												

There are 1497 specimen_ids each have 35 or 80 rows.

One more join to do of `meta` and `abdata` to associate all the metadata about the individual and their race, biological sex, and infancy vaccination status together with Antibody levels...

```
ab <- inner_join(abdata, meta)
```

Joining with ``by = join_by(specimen_id)``

```
head(ab)
```

	specimen_id	isotype	is_antigen_specific	antigen	MFI	MFI_normalised
1	1	IgE	FALSE	Total	1110.21154	2.493425
2	1	IgE	FALSE	Total	2708.91616	2.493425
3	1	IgG	TRUE	PT	68.56614	3.736992
4	1	IgG	TRUE	PRN	332.12718	2.602350
5	1	IgG	TRUE	FHA	1887.12263	34.050956
6	1	IgE	TRUE	ACT	0.10000	1.000000

	unit	lower_limit_of_detection	subject_id	infancy_vac	biological_sex
1	UG/ML	2.096133	1	wP	Female
2	IU/ML	29.170000	1	wP	Female
3	IU/ML	0.530000	1	wP	Female
4	IU/ML	6.205949	1	wP	Female
5	IU/ML	4.679535	1	wP	Female
6	IU/ML	2.816431	1	wP	Female

	ethnicity	race	year_of_birth	date_of_boost	dataset
1	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
2	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
3	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
4	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
5	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
6	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset

	actual_day_relative_to_boost	planned_day_relative_to_boost	specimen_type
1	-3		Blood
2	-3		Blood
3	-3		Blood
4	-3		Blood
5	-3		Blood
6	-3		Blood

	visit
1	1
2	1
3	1
4	1
5	1
6	1

Q. How many Ab measurements do we have?

```
nrow(ab)
```

```
[1] 52576
```

How many isotypes

```
table(ab$isotype)
```

```
  IgE   IgG  IgG1  IgG2  IgG3  IgG4
6698  5389 10117 10124 10124 10124
```

How many antigens?

```
table(ab$antigen)
```

```
  ACT  BETV1    DT  FELD1    FHA  FIM2/3  LOLP1    LOS Measles    OVA
1970   1970  4978   1970   5372   4978   1970   1970   1970   4978
  PD1    PRN    PT    PTM  Total    TT
1970   5372  5372   1970   788   4978
```

Let's focus in on IgG - one of the main antibody types responsive to bacteria or virial infections

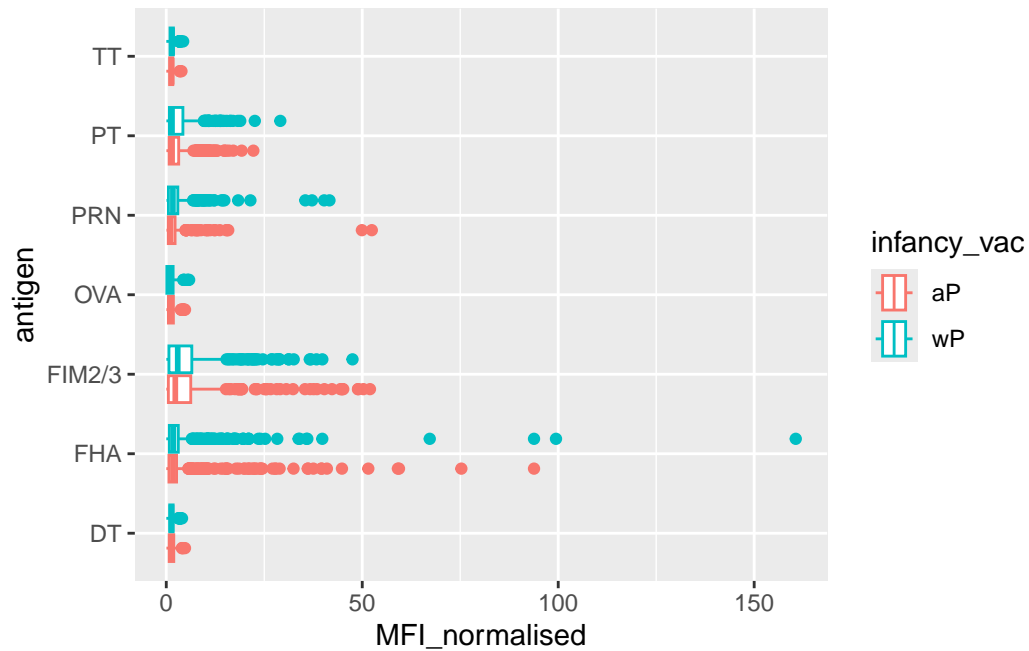
```
igg <- filter(ab, isotype == "IgG")
head(igg)
```

```
  specimen_id isotype is_antigen_specific antigen      MFI MFI_normalised
1           1    IgG                TRUE      PT   68.56614         3.736992
2           1    IgG                TRUE      PRN  332.12718         2.602350
3           1    IgG                TRUE      FHA 1887.12263        34.050956
4          19    IgG                TRUE      PT   20.11607         1.096366
5          19    IgG                TRUE      PRN  976.67419         7.652635
6          19    IgG                TRUE      FHA   60.76626         1.096457
  unit lower_limit_of_detection subject_id infancy_vac biological_sex
1 IU/ML                0.530000          1          wP          Female
2 IU/ML                6.205949          1          wP          Female
3 IU/ML                4.679535          1          wP          Female
```

4	IU/ML	0.530000	3	wP	Female
5	IU/ML	6.205949	3	wP	Female
6	IU/ML	4.679535	3	wP	Female
	ethnicity	race	year_of_birth	date_of_boost	dataset
1	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
2	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
3	Not Hispanic or Latino	White	1986-01-01	2016-09-12	2020_dataset
4	Unknown	White	1983-01-01	2016-10-10	2020_dataset
5	Unknown	White	1983-01-01	2016-10-10	2020_dataset
6	Unknown	White	1983-01-01	2016-10-10	2020_dataset
	actual_day_relative_to_boost	planned_day_relative_to_boost	specimen_type		
1	-3		0	Blood	
2	-3		0	Blood	
3	-3		0	Blood	
4	-3		0	Blood	
5	-3		0	Blood	
6	-3		0	Blood	
	visit				
1	1				
2	1				
3	1				
4	1				
5	1				
6	1				

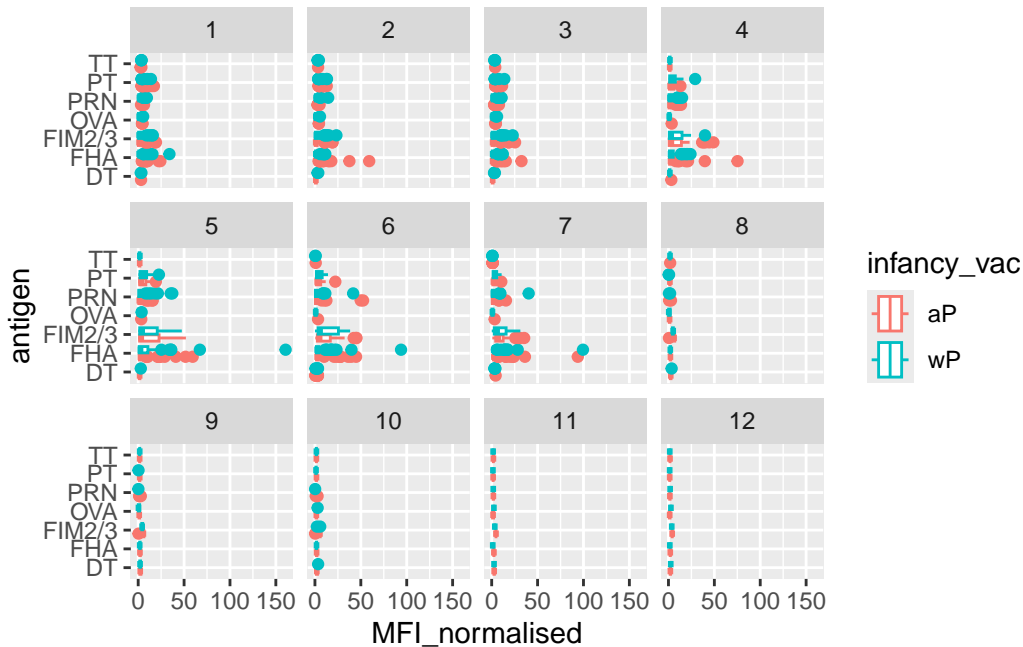
Make a first plot of MFI (Mean Fluorescence Intensity - a measure of how much is detected) for each antigen.

```
ggplot(igg, aes(MFI_normalised, antigen, col = infancy_vac)) +
  geom_boxplot()
```



Q13. Complete the following code to make a summary boxplot of Ab titer levels for all antigens:

```
ggplot(igg, aes(MFI_normalised, antigen, col = infancy_vac)) +
  geom_boxplot() +
  facet_wrap(~visit)
```

Q14. What antigens show differences in the level of IgG1 antibody titers recognizing them over time? Why these and not others?

Antigens PT, PRN, FIM2/3, and FHA show differences in the level of IgG antibody titers (MFI) over time. The Terminology search function is still under development. But, I can guess that these four antigens are more influential to the main pertussis toxin.

```
table(igg$visit)
```

```

 1    2    3    4    5    6    7    8    9   10   11   12
902 902 930 559 559 540 525 150 147 133  21  21

```

Exclude visits 8~12. Focus on the first 7 visits. Looks like we don't have data yet for all subjects in terms of visits 8 onwards. So let's exclude these.

```

igg_7 <- filter(igg, visit %in% 1:7)
table(igg_7$visit)

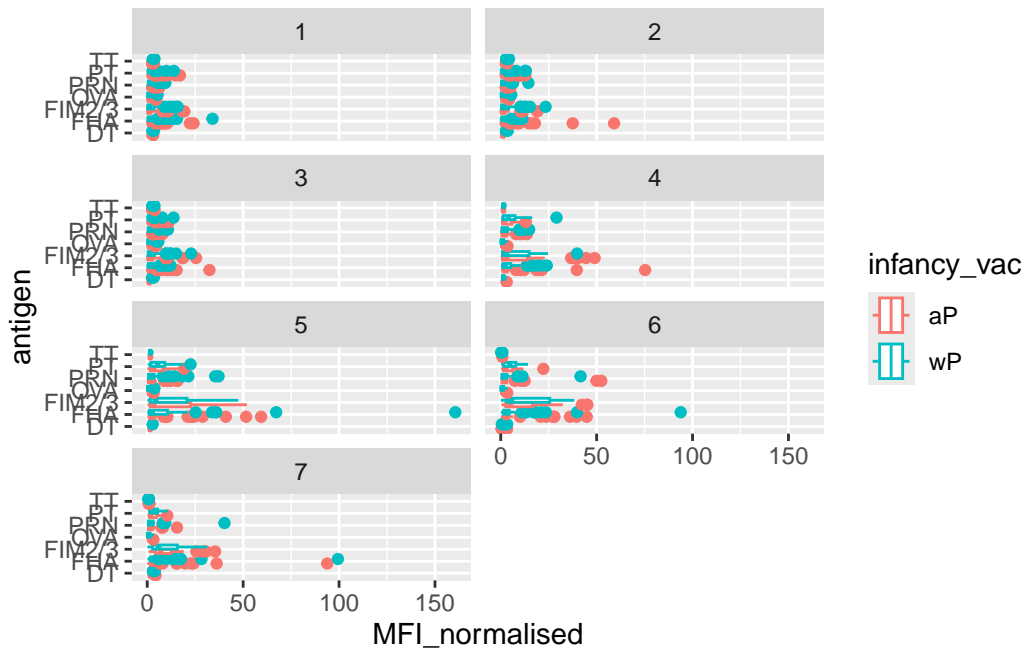
```

```

 1    2    3    4    5    6    7
902 902 930 559 559 540 525

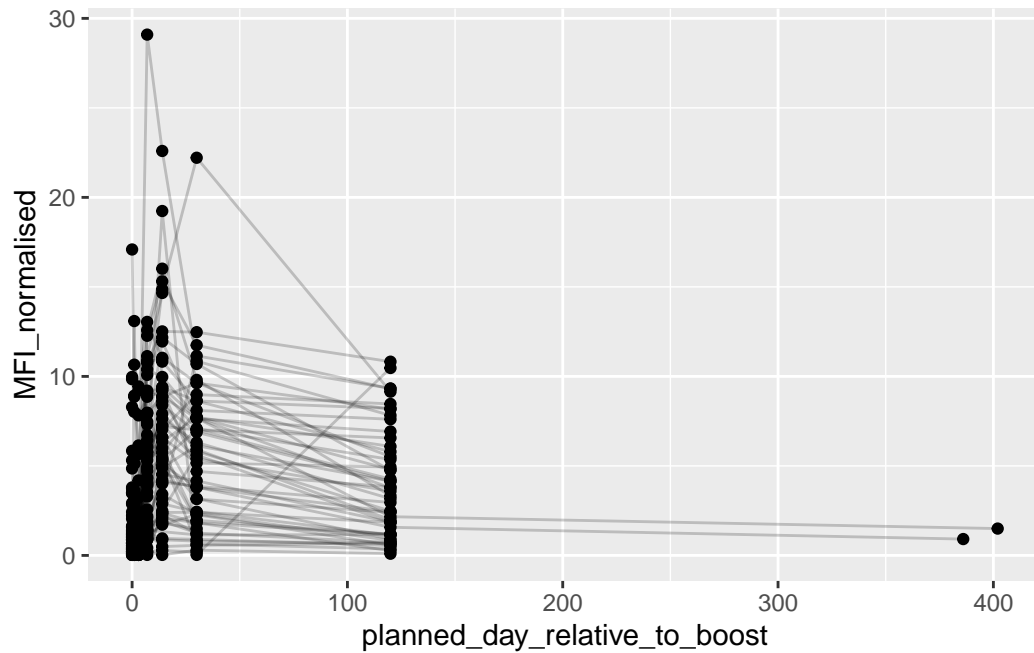
```

```
ggplot(igg_7, aes(MFI_normalised, antigen, col = infancy_vac)) +
  geom_boxplot() +
  facet_wrap(~visit
            , ncol = 2)
```



Let's try a different plot. First focus on one antigen, start with PT (Pertussis Toxin) and plot visit or time on the x-axis and MFI_normalised on the y-axis.

```
igg_7 %>%
  filter(antigen == "PT", dataset == "2020_dataset") %>%
  ggplot() +
  aes(planned_day_relative_to_boost, MFI_normalised, group = subject_id) +
  geom_point() +
  geom_line(alpha = 0.2)
```



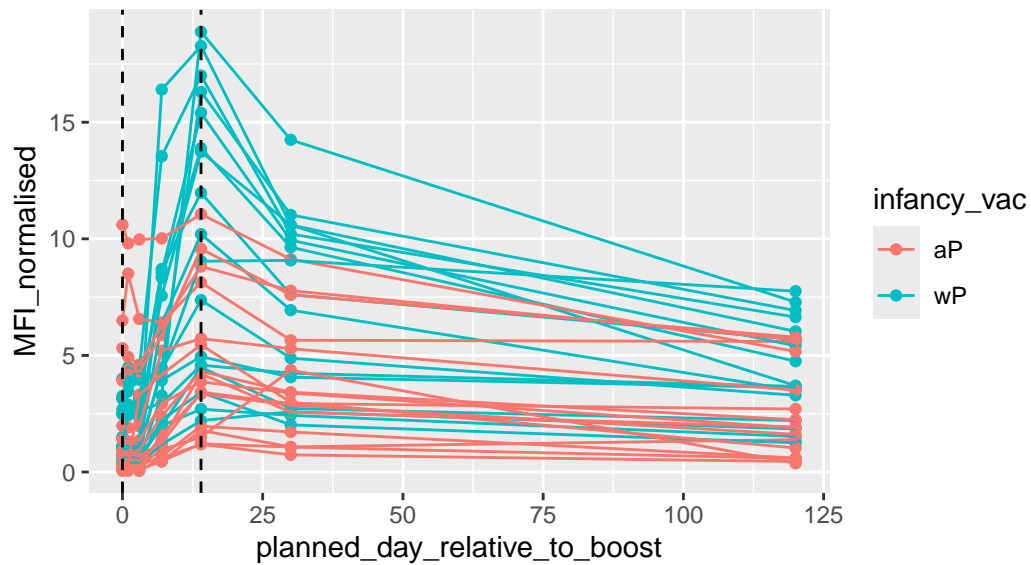
Must filter to a single year.

```
abdata.21 <- ab %>% filter(dataset == "2021_dataset")

abdata.21 %>%
  filter(isotype == "IgG", antigen == "PT") %>%
  ggplot() +
    aes(x=planned_day_relative_to_boost,
        y=MFI_normalised,
        col=infancy_vac,
        group=subject_id) +
    geom_point() +
    geom_line() +
    geom_vline(xintercept=0, linetype="dashed") +
    geom_vline(xintercept=14, linetype="dashed") +
    labs(title="2021 dataset IgG PT",
         subtitle = "Dashed lines indicate day 0 (pre-boost) and 14 (apparent peak levels)")
```

2021 dataset IgG PT

Dashed lines indicate day 0 (pre-boost) and 14 (apparent peak levels)



Let's finish here for today. We are beginning to see some interesting differences between aP and wP individuals. There is likely lots of other interesting things to find in this dataset...

```
library(lubridate)
```