Compare new and old 2021 data Public Data runs

Emily Markowitz (emily.markowitz@nodiff_between_cols.gov)

2021 old public data: Load and compile data

column names have been changed to match with the new data column names

```
FALSE       year           srvy                                survey            haul              cruise
FALSE  Min.   :1982   AI     :141763   Aleutian Islands       :141763   Min.   :     1.0   Min.   :   14
FALSE  1st Qu.:1996   BSSLOPE: 47448   Bering Sea Slope       : 47448   1st Qu.:    57.0   1st Qu.:19960
FALSE  Median :2005   EBS    :460047   Gulf of Alaska         :252849   Median :   112.0   Median :20050
FALSE  Mean   :2004   GOA    :252849   Northern Bering Sea    : 38773   Mean   :   118.2   Mean   :20039
FALSE  3rd Qu.:2013   NBS    : 38773   Southeastern Bering Sea:460047   3rd Qu.:   168.0   3rd Qu.:20130
FALSE  Max.   :2021                                                     Max.   :201801.0   Max.   :20210
FALSE
FALSE                common_name      species_code                    scientific_name    longitude_dd
FALSE  walleye pollock    : 27674   Min.   :    1   Gadus chalcogrammus      : 27674   Min.   :-180.0
FALSE  Pacific cod        : 26431   1st Qu.:21110   Gadus macrocephalus      : 26431   1st Qu.:-171.5
FALSE  arrowtooth flounder: 23021   Median :41201   Atheresthes stomias      : 23021   Median :-166.5
FALSE  Pacific halibut    : 22617   Mean   :46684   Hippoglossus stenolepis  : 22617   Mean   :-143.6
FALSE  flathead sole      : 20010   3rd Qu.:72500   Hippoglossoides elassodon: 20010   3rd Qu.:-157.2
FALSE  (Other)            :716356   Max.   :99998   (Other)                  :785474   Max.   : 180.0
FALSE  NA's               :104771   NA's   :5       NA's                     : 35653
FALSE   latitude_dd       stratum         station         cpue_kgha          cpue_noha        surface_te
FALSE  Min.   :51.19   Min.   :  0.0   Q-23   :  1672   Min.   :    0.00   Min.   :-9999.00   Min.   :-9
FALSE  1st Qu.:55.27   1st Qu.: 31.0   N-24   :  1634   1st Qu.:    0.05   1st Qu.:-9999.00   1st Qu.:
FALSE  Median :57.34   Median : 50.0   H-01   :  1626   Median :    0.42   Median :-9999.00   Median :
FALSE  Mean   :57.09   Mean   :126.8   G-02   :  1623   Mean   :   11.52   Mean   :-5274.57   Mean   : -
FALSE  3rd Qu.:59.01   3rd Qu.:134.0   ON2524 :  1602   3rd Qu.:    3.25   3rd Qu.:    1.96   3rd Qu.:
FALSE  Max.   :65.34   Max.   :794.0   (Other):901042   Max.   :32262.35   Max.   :44817.02   Max.   :
FALSE                  NA's   :9202    NA's   : 31681                      NA's   :1130
FALSE  bottom_temperature_c    depth_m          vessel_id       vessel                  date
FALSE  Min.   :-9999.0      Min.   :    5.0   Min.   :  1.0   Length:940880      06/25/2012:  1021
FALSE  1st Qu.:    2.0      1st Qu.:   68.0   1st Qu.: 88.0   Class :character   07/10/2010:  1016
FALSE  Median :    3.8      Median :  101.0   Median : 94.0   Mode  :character   07/07/2012:   990
FALSE  Mean   : -516.2      Mean   :  144.4   Mean   :111.3                      07/11/2010:   989
FALSE  3rd Qu.:    4.9      3rd Qu.:  160.0   3rd Qu.:143.0                      07/26/2021:   924
FALSE  Max.   :  311.0      Max.   :91264.0   Max.   :560.0                      06/25/2010:   921
FALSE                                                                          (Other)   :935019
```

2021 new public data: Load and compile data

```
FALSE       year           srvy
FALSE  Min.   :1982   AI     :141877
```

```
FALSE  1st Qu.:1996   BSSlope: 47406
FALSE  Median :2005   EBS    :460325
FALSE  Mean   :2004   GOA    :252965
FALSE  3rd Qu.:2013   NBS    : 20179
FALSE  Max.   :2021
FALSE
FALSE                                                                                  survey
FALSE  Aleutian Islands Bottom Trawl Survey                                               :141877
FALSE  Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey                             :460325
FALSE  Eastern Bering Sea Slope Bottom Trawl Survey                                       : 47406
FALSE  Gulf of Alaska Bottom Trawl Survey                                                 :252965
FALSE  Northern Bering Sea Crab/Groundfish Survey - Eastern Bering Sea Shelf Survey Extension: 20179
FALSE
FALSE
FALSE    survey_id         cruise          stratum         station            haul       vessel_name
FALSE  Min.   : 47.0   Min.   :198201   Min.   : 10.0   Q-23   :  1677   Min.   :  1.0   Length:922752
FALSE  1st Qu.: 47.0   1st Qu.:199601   1st Qu.: 31.0   N-24   :  1636   1st Qu.: 57.0   Class :charact
FALSE  Median : 98.0   Median :200501   Median : 50.0   H-01   :  1626   Median :112.0   Mode  :charact
FALSE  Mean   : 76.9   Mean   :200403   Mean   :127.5   G-02   :  1624   Mean   :116.8
FALSE  3rd Qu.: 98.0   3rd Qu.:201301   3rd Qu.:134.0   ON2524 :  1602   3rd Qu.:168.0
FALSE  Max.   :143.0   Max.   :202102   Max.   :794.0   G-01   :  1601   Max.   :454.0
FALSE                                                  (Other):912986
FALSE    vessel_id           date         latitude_dd     longitude_dd     species_code   taxon_confider
FALSE  Min.   :  1   06/25/2012:  1021   Min.   :51.19   Min.   :-180.0   Min.   :    1   Length:922752
FALSE  1st Qu.: 88   07/10/2010:  1016   1st Qu.:55.18   1st Qu.:-171.6   1st Qu.:21110   Class :charact
FALSE  Median : 94   07/07/2012:   991   Median :57.33   Median :-166.5   Median :41201   Mode  :charact
FALSE  Mean   :112   07/11/2010:   989   Mean   :56.96   Mean   :-143.2   Mean   :46633
FALSE  3rd Qu.:143   06/25/2010:   921   3rd Qu.:58.99   3rd Qu.:-156.8   3rd Qu.:72500
FALSE  Max.   :560   07/18/2012:   916   Max.   :65.34   Max.   : 180.0   Max.   :99999
FALSE                (Other)   :916898
FALSE               common_name                          scientific_name     cpue_kgha          cpue_kgkm2
FALSE  walleye pollock    : 27290   Gadus chalcogrammus      : 27290   Min.   :    0.00   Min.   :
FALSE  Pacific cod        : 26163   Gadus macrocephalus      : 26163   1st Qu.:    0.05   1st Qu.:
FALSE  arrowtooth flounder: 23002   Atheresthes stomias      : 23002   Median :    0.42   Median :     4
FALSE  Pacific halibut    : 22478   Hippoglossus stenolepis  : 22478   Mean   :   11.67   Mean   :   110
FALSE  flathead sole      : 19997   Hippoglossoides elassodon: 19997   3rd Qu.:    3.29   3rd Qu.:    31
FALSE  (Other)            :702953   (Other)                  :768463   Max.   :32262.35   Max.   :322623
FALSE  NA's               :100869   NA's                     : 35359
FALSE  cpue_kg1000km2        cpue_noha           cpue_nokm2         weight_kg            count
FALSE  Min.   :1.500e+01   Min.   :     0.04   Min.   :       4   Min.   :    0.00   Min.   :     0.0
FALSE  1st Qu.:4.982e+03   1st Qu.:     0.53   1st Qu.:      53   1st Qu.:    0.18   1st Qu.:     1.0
FALSE  Median :4.201e+04   Median :     2.01   Median :     201   Median :    1.59   Median :     5.0
FALSE  Mean   :1.167e+06   Mean   :    42.53   Mean   :    4253   Mean   :   41.40   Mean   :   148.3
FALSE  3rd Qu.:3.289e+05   3rd Qu.:    10.50   3rd Qu.:    1050   3rd Qu.:   12.46   3rd Qu.:    31.0
FALSE  Max.   :3.226e+09   Max.   :217807.80   Max.   :21780780   Max.   :42200.15   Max.   :867119.0
FALSE                      NA's   :113703      NA's   :113703
FALSE     depth_m       bottom_temperature_c surface_temperature_c distance_fished_km  net_width_m
FALSE  Min.   :   9.0   Min.   :-2.10        Min.   :-1.100        Min.   :0.135      Min.   : 7.51
FALSE  1st Qu.:  69.0   1st Qu.: 2.40        1st Qu.: 5.700        1st Qu.:1.523      1st Qu.:15.69
FALSE  Median : 103.0   Median : 3.80        Median : 7.400        Median :2.683      Median :16.52
FALSE  Mean   : 146.2   Mean   : 3.61        Mean   : 7.734        Mean   :2.313      Mean   :16.65
FALSE  3rd Qu.: 162.0   3rd Qu.: 4.90        3rd Qu.: 9.300        3rd Qu.:2.844      3rd Qu.:17.32
FALSE  Max.   :1200.0   Max.   :15.30        Max.   :18.100        Max.   :8.241      Max.   :30.90
FALSE                   NA's   :48632        NA's   :19709
```

```
FALSE  area_swept_ha        hauljoin         cruisejoin
FALSE  Min.   : 0.2314   Min.   : -21018   Min.   :   -752
FALSE  1st Qu.: 2.4905   1st Qu.: -11924   1st Qu.:   -682
FALSE  Median : 4.2653   Median :  -3453   Median :   -610
FALSE  Mean   : 3.8937   Mean   : 295754   Mean   : 298825
FALSE  3rd Qu.: 4.8103   3rd Qu.: 821815   3rd Qu.: 837800
FALSE  Max.   :22.2721   Max.   :1225635   Max.   :1225395
FALSE
```

Bind new and old 2021 data to see where differences are

- -9999 –> NA from cpue_noha and cpue_kgha in the old data so we can compare new and old more easily
- commas (",") removed from old data (e.g., commas in "Clark, 2006" vs "Clark 2006") because I bet it doesn't matter. **Should the comma stay for the final product?**

```
dat <- dplyr::full_join(
  x = data_new %>%
    dplyr::mutate(dataset = "new",
                  scientific_name = gsub(pattern = ",", replacement = "",
                                         x = scientific_name, fixed = TRUE)),
  y = data_old %>%
    dplyr::mutate(dataset = "old",
                  cpue_kgha = ifelse(cpue_kgha == -9999, NA, cpue_kgha),
                  cpue_noha = ifelse(cpue_noha == -9999, NA, cpue_noha)),
  by = c("year", "vessel_id", "stratum", "station", "species_code", "haul", "cruise",  "srvy"))
dat
```

```
FALSE # A tibble: 988,853 x 46
FALSE     year srvy  survey.x        survey_id cruise stratum station  haul vessel_name vessel_id date.
FALSE    <dbl> <chr> <chr>               <dbl>  <dbl>   <dbl> <chr>   <dbl> <chr>           <dbl> <chr>
FALSE  1  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  2  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  3  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  4  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  5  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  6  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  7  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  8  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE  9  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE 10  2002 AI    Aleutian Islan~        52 200201     722 303-64      2 F/V Vester~        94 05/17
FALSE # ... with 988,843 more rows, and 34 more variables: longitude_dd.x <dbl>, species_code <dbl>,
FALSE #   taxon_confidence <chr>, common_name.x <chr>, scientific_name.x <chr>, cpue_kgha.x <dbl>, cpue
FALSE #   cpue_kg1000km2 <dbl>, cpue_noha.x <dbl>, cpue_nokm2 <dbl>, weight_kg <dbl>, count <dbl>, depth
FALSE #   bottom_temperature_c.x <dbl>, surface_temperature_c.x <dbl>, distance_fished_km <dbl>, net_wi
FALSE #   area_swept_ha <dbl>, hauljoin <dbl>, cruisejoin <dbl>, dataset.x <chr>, survey.y <chr>,
FALSE #   common_name.y <chr>, scientific_name.y <chr>, longitude_dd.y <dbl>, latitude_dd.y <dbl>, cpue
FALSE #   cpue_noha.y <dbl>, surface_temperature_c.y <dbl>, bottom_temperature_c.y <dbl>, depth_m.y <dbl
```

What observations are in new and not in old, visa versa

column dataset = - both = observations are in both new and old datasets - old = observations are only in the old dataset - new = observations are only in the new dataset

```
dat0 <- dat %>%
  dplyr::mutate(dataset = dplyr::case_when(
    (dataset.x == "new" & dataset.y == "old") ~ "both",
    dataset.y == "old" ~ "old",
    dataset.x == "new" ~ "new",
    TRUE ~ "neither"))
dat0 %>%
  dplyr::select(dataset) %>%
  table()
```

```
FALSE .
FALSE   both    new    old
FALSE 879366  47973  61514
```

datasets only in the new data

year by survey

```
dat0 %>%
  dplyr::filter(dataset == "new") %>% # data only in new data, not in old data
  dplyr::select(year, srvy) %>%
  table()
```

```
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA  NBS
FALSE   1986    0       0    8    0    0
FALSE   1987    0       0    5    3    0
FALSE   1988    0       0    7    0    0
FALSE   1989    0       0   57    0    0
FALSE   1990    0       0   17   42    0
FALSE   1991   61       0   22    0    0
FALSE   1992    0       0    5    0    0
FALSE   1993    0       0   17   39    0
FALSE   1994   30       0    1    0    0
FALSE   1995    0       0    2    0    0
FALSE   1996    0       0   18   10    0
FALSE   1997    2       0    8    0    0
FALSE   1999    0       0    2    0    0
FALSE   2000    0       0    1    0    0
FALSE   2002    0    6198    2    0    0
FALSE   2004    0    9602    3    0    0
FALSE   2005    0       0    2    2    0
FALSE   2006    0       0   10    0    0
FALSE   2007    0       0    2    0    0
FALSE   2008    0    8262   13    0    0
FALSE   2009    0       0   10    0    0
```

```
FALSE   2010   11    8531    4    0    2
FALSE   2011    0       0    7    0    0
FALSE   2012    0    7477    2    0    0
FALSE   2013    0       0   34    1    0
FALSE   2014   10       0    4    0    0
FALSE   2015    0       0    3    1    0
FALSE   2016    5    7336    0    0    0
FALSE   2017    0       0    5    4    1
FALSE   2018    2       0    0    0    0
FALSE   2019    0       0    6   14    5
FALSE   2021    0       0    3   35    7
```

species_code by survey

```
dat0 %>%
  dplyr::filter(dataset == "new") %>% # data only in new data, not in old data
  dplyr::select(species_code, srvy) %>%
  table()
```

```
FALSE             srvy
FALSE species_code  AI BSSlope EBS GOA NBS
FALSE        1       0      12   0   0   0
FALSE        10      0      29   0   0   0
FALSE        21      0     254   0   0   0
FALSE        160     0       0   0   1   0
FALSE        310     0       1   0   0   0
FALSE        320     0     107   0   0   0
FALSE        400     0       1   0   0   0
FALSE        401     0      57   0   0   0
FALSE        402     0     134   0   0   0
FALSE        405     0      13   0   0   0
FALSE        410     0      12   0   0   0
FALSE        411     0      13   0   0   0
FALSE        435     0     509   0   0   0
FALSE        436     0     153   0   0   0
FALSE        440     0       1   0   0   0
FALSE        455     0     211   0   0   0
FALSE        456     0      21   0   0   0
FALSE        460     0     269   0   0   0
FALSE        461     0     106   0   0   0
FALSE        471     0     168   0   0   0
FALSE        472     0     886   0   0   0
FALSE        473     0      78   0   0   0
FALSE        474     0     131   0   0   0
FALSE        475     0     392   0   0   0
FALSE        476     0       7   0   0   0
FALSE        480     0     323   0   0   0
FALSE        481     0       8   0   0   0
FALSE        485     0     514   0   0   0
FALSE        486     0      65   0   0   0
FALSE        495     0       4   0   0   0
FALSE        10100   0       0   0   1   0
```

FALSE	10110	0	703	0	0	0
FALSE	10112	0	967	0	0	0
FALSE	10115	0	795	0	0	0
FALSE	10120	0	330	0	0	0
FALSE	10130	0	564	0	0	0
FALSE	10150	0	3	0	0	0
FALSE	10180	0	159	0	0	0
FALSE	10190	0	120	0	0	0
FALSE	10200	0	557	0	0	0
FALSE	10261	0	48	0	0	0
FALSE	20002	0	2	0	0	0
FALSE	20006	0	307	0	0	0
FALSE	20034	0	1	0	0	0
FALSE	20035	0	7	0	0	0
FALSE	20038	0	572	0	0	0
FALSE	20039	0	0	0	0	1
FALSE	20050	0	4	0	0	0
FALSE	20100	0	1	0	0	0
FALSE	20122	0	2	0	0	0
FALSE	20130	0	2	0	0	0
FALSE	20510	0	593	0	0	0
FALSE	20572	0	1	0	0	0
FALSE	20615	0	137	0	0	0
FALSE	20619	0	209	0	0	0
FALSE	20620	0	63	0	0	0
FALSE	20622	0	636	0	0	0
FALSE	20624	0	34	0	0	0
FALSE	20660	0	3	0	0	0
FALSE	20720	0	50	0	0	0
FALSE	20916	0	1	0	0	0
FALSE	21010	0	181	0	0	0
FALSE	21110	0	12	0	0	0
FALSE	21210	0	1	0	0	0
FALSE	21220	0	228	0	0	0
FALSE	21230	0	789	0	0	0
FALSE	21232	0	552	0	0	0
FALSE	21238	0	2	0	0	0
FALSE	21239	0	15	0	0	0
FALSE	21305	0	62	0	0	0
FALSE	21316	0	1	0	0	0
FALSE	21341	0	548	0	0	0
FALSE	21347	0	14	0	0	0
FALSE	21348	0	1	0	0	0
FALSE	21354	0	100	0	0	0
FALSE	21370	0	13	0	0	0
FALSE	21390	0	412	0	0	0
FALSE	21395	0	180	0	0	0
FALSE	21404	0	0	0	0	2
FALSE	21420	0	344	0	0	0
FALSE	21422	0	7	0	0	0
FALSE	21438	0	152	0	0	0
FALSE	21439	0	189	0	0	0
FALSE	21440	0	61	0	0	0
FALSE	21441	0	2	0	0	0

FALSE	21444	0	8	0	0	0
FALSE	21720	0	270	0	0	0
FALSE	21727	0	28	0	0	0
FALSE	21731	0	141	0	0	0
FALSE	21740	0	495	0	0	0
FALSE	21800	0	12	0	0	0
FALSE	21805	0	14	0	0	0
FALSE	21810	0	8	0	0	0
FALSE	21811	0	1	0	0	0
FALSE	21820	0	0	0	1	0
FALSE	21921	0	17	0	0	0
FALSE	22010	0	5	0	0	0
FALSE	22175	0	234	0	0	0
FALSE	22200	0	155	0	0	0
FALSE	22206	0	7	0	0	0
FALSE	22214	0	7	0	0	0
FALSE	22215	0	339	0	0	0
FALSE	22216	0	178	0	0	0
FALSE	22217	0	1	0	0	0
FALSE	22219	0	24	0	0	0
FALSE	22220	0	349	0	0	0
FALSE	22222	0	6	0	0	0
FALSE	22223	0	33	0	0	0
FALSE	22226	0	1	0	0	0
FALSE	22228	0	198	0	0	0
FALSE	22229	0	288	0	0	0
FALSE	22231	0	34	0	0	0
FALSE	22232	0	2	0	0	0
FALSE	22233	0	22	0	0	0
FALSE	22234	0	7	0	0	0
FALSE	22235	0	162	0	0	0
FALSE	22236	0	181	0	0	0
FALSE	22239	0	1	0	0	0
FALSE	22240	0	1	0	0	0
FALSE	22241	0	31	0	0	0
FALSE	22244	0	24	0	0	0
FALSE	22245	0	10	0	0	0
FALSE	22246	0	2	0	0	0
FALSE	22249	0	32	0	0	0
FALSE	22250	0	26	0	0	0
FALSE	22252	0	2	0	0	0
FALSE	22253	0	7	0	0	0
FALSE	22254	0	0	0	0	1
FALSE	22255	0	2	0	0	0
FALSE	22260	0	26	0	0	0
FALSE	22261	0	6	0	0	0
FALSE	22262	0	49	0	0	0
FALSE	22271	0	5	0	0	0
FALSE	22276	0	13	0	0	0
FALSE	22277	0	2	0	0	0
FALSE	22281	0	3	0	0	0
FALSE	22282	0	7	0	0	0
FALSE	22283	0	2	0	0	0
FALSE	22300	0	6	0	0	0

FALSE	22310	0	33	0	0	0
FALSE	22320	0	34	0	0	0
FALSE	22390	0	1	0	0	0
FALSE	22410	0	7	0	0	0
FALSE	22420	0	24	0	0	0
FALSE	22461	0	1	0	0	0
FALSE	22500	0	1	0	0	0
FALSE	22570	0	1	0	0	0
FALSE	22600	0	74	0	0	0
FALSE	22601	0	376	0	0	0
FALSE	22602	0	319	0	0	0
FALSE	22603	0	70	0	0	0
FALSE	22610	0	34	0	0	0
FALSE	22620	0	10	0	0	0
FALSE	22623	0	5	0	0	0
FALSE	22624	0	15	0	0	0
FALSE	22625	0	155	0	0	0
FALSE	22648	0	5	0	0	0
FALSE	22649	0	1	0	0	0
FALSE	22701	0	1	0	0	0
FALSE	22704	0	6	0	0	0
FALSE	22815	0	2	0	0	0
FALSE	22900	0	14	0	0	0
FALSE	22912	0	34	0	0	0
FALSE	22914	0	32	0	0	0
FALSE	22915	0	71	0	0	0
FALSE	22951	0	2	0	0	0
FALSE	23010	0	89	0	0	0
FALSE	23220	0	1	0	0	0
FALSE	23230	0	3	0	0	0
FALSE	23235	0	22	0	0	0
FALSE	23240	0	1	0	0	0
FALSE	23602	0	1	0	0	0
FALSE	23603	0	20	0	0	0
FALSE	23620	0	2	0	0	0
FALSE	23622	0	3	0	0	0
FALSE	23657	0	10	0	0	0
FALSE	23710	0	5	0	0	0
FALSE	23836	0	33	0	0	0
FALSE	23850	0	1	0	0	0
FALSE	23862	0	2	0	0	0
FALSE	23962	0	19	0	0	0
FALSE	24001	0	41	0	0	0
FALSE	24109	0	6	0	0	0
FALSE	24110	0	329	0	0	0
FALSE	24111	0	23	0	0	0
FALSE	24112	0	360	0	0	0
FALSE	24114	0	34	0	0	0
FALSE	24115	0	2	0	0	0
FALSE	24130	0	45	0	0	0
FALSE	24150	0	3	0	0	0
FALSE	24152	0	21	0	0	0
FALSE	24153	0	6	0	0	0
FALSE	24158	0	2	0	0	0

```
FALSE        24160   0      19   0   0   0
FALSE        24166   0       1   0   0   0
FALSE        24180   0       5   0   0   0
FALSE        24185   0       1   0   0   0
FALSE        24187   0     295   0   0   0
FALSE        24191   0      24   0   0   0
FALSE        24195   0     538   0   0   0
FALSE  [ reached getOption("max.print") -- omitted 594 rows ]
```

datasets only in the old data

year by survey

```
dat0 %>%
  dplyr::filter(dataset == "old") %>% # data only in old data, not in new data
  dplyr::select(year, srvy) %>%
  table()
```

```
FALSE       srvy
FALSE year     AI BSSLOPE  EBS  GOA  NBS
FALSE   1982    0       0    2    0 1976
FALSE   1985    0       0    0    0 4156
FALSE   1987    0       0    0   33    0
FALSE   1988    0       0    0    0 3728
FALSE   1991    0       0    0    0 2488
FALSE   2002    0    6198    0    0    0
FALSE   2004    0    9661    0    0    0
FALSE   2008    0    8255    0    0    0
FALSE   2010    0    8531    0    0    0
FALSE   2012    0    7474    0    0    0
FALSE   2016    4    7329    0    0    0
FALSE   2018    1       0    0    0 1678
```

species_code by survey

```
dat0 %>%
  dplyr::filter(dataset == "old") %>% # data only in old data, not in new data
  dplyr::select(species_code, srvy) %>%
  table()
```

```
FALSE             srvy
FALSE species_code  AI BSSLOPE EBS GOA NBS
FALSE        1       0      12   0   0   1
FALSE        10      0      29   0   0   0
FALSE        21      0     254   0   0   0
FALSE        310     0       1   0   1   0
FALSE        320     0     107   0   0   0
FALSE        400     0       1   0   0   4
FALSE        401     0      57   0   0   1
```

FALSE	402	0	134	0	0	0
FALSE	405	0	13	0	0	0
FALSE	410	0	12	0	0	0
FALSE	411	0	13	0	0	0
FALSE	435	0	509	0	0	0
FALSE	436	0	153	0	0	0
FALSE	440	0	1	0	0	0
FALSE	450	0	0	0	0	1
FALSE	455	0	212	0	0	0
FALSE	456	0	21	0	0	0
FALSE	460	0	269	0	0	0
FALSE	461	0	106	0	0	0
FALSE	471	0	168	0	0	39
FALSE	472	0	887	0	0	0
FALSE	473	0	78	0	0	0
FALSE	474	0	131	0	0	4
FALSE	475	0	393	0	0	0
FALSE	476	0	7	0	0	0
FALSE	480	0	324	0	0	0
FALSE	481	0	8	0	0	0
FALSE	485	0	515	0	0	0
FALSE	486	0	66	0	0	0
FALSE	495	0	4	0	0	0
FALSE	10001	0	0	0	0	1
FALSE	10110	0	704	0	1	8
FALSE	10112	0	968	0	0	5
FALSE	10115	0	796	0	0	17
FALSE	10120	0	330	0	1	60
FALSE	10130	0	564	0	1	10
FALSE	10140	0	0	0	0	137
FALSE	10150	0	3	0	0	0
FALSE	10155	0	0	0	0	22
FALSE	10170	0	0	0	1	0
FALSE	10180	0	159	0	1	0
FALSE	10190	0	120	0	0	0
FALSE	10200	0	558	0	0	0
FALSE	10210	0	0	0	1	387
FALSE	10211	0	0	0	0	120
FALSE	10212	0	0	0	0	53
FALSE	10220	0	0	0	1	260
FALSE	10260	0	0	0	1	69
FALSE	10261	0	48	0	0	46
FALSE	10270	0	0	0	1	0
FALSE	10285	0	0	0	1	372
FALSE	20000	0	0	0	0	5
FALSE	20001	0	0	0	0	18
FALSE	20002	0	2	0	0	0
FALSE	20006	0	307	0	0	11
FALSE	20007	0	0	0	0	1
FALSE	20034	0	1	0	0	0
FALSE	20035	0	7	0	0	0
FALSE	20038	0	572	0	0	0
FALSE	20040	0	0	0	1	228
FALSE	20041	0	0	0	0	15

FALSE	20050	0	4	0	0	1
FALSE	20051	0	0	0	0	15
FALSE	20061	0	0	0	0	56
FALSE	20100	0	1	0	0	0
FALSE	20122	0	2	0	0	0
FALSE	20130	0	2	0	0	0
FALSE	20202	0	0	0	0	11
FALSE	20322	0	0	0	0	22
FALSE	20510	0	593	0	1	0
FALSE	20572	0	1	0	0	0
FALSE	20615	0	137	0	0	0
FALSE	20619	0	209	0	0	0
FALSE	20620	0	63	0	0	0
FALSE	20622	0	637	0	0	0
FALSE	20624	0	34	0	0	0
FALSE	20660	0	3	0	0	0
FALSE	20700	0	0	0	0	1
FALSE	20720	0	50	0	0	0
FALSE	20916	0	1	0	0	0
FALSE	21010	0	181	0	0	0
FALSE	21110	0	12	0	0	168
FALSE	21210	0	1	0	0	0
FALSE	21220	0	228	0	0	0
FALSE	21230	0	790	0	0	0
FALSE	21232	0	552	0	0	0
FALSE	21238	0	2	0	0	0
FALSE	21239	0	15	0	0	0
FALSE	21300	0	0	0	0	16
FALSE	21305	0	62	0	0	0
FALSE	21313	0	0	0	0	52
FALSE	21314	0	0	0	0	91
FALSE	21315	0	0	0	0	139
FALSE	21316	0	1	0	0	8
FALSE	21331	0	0	0	0	2
FALSE	21332	0	0	0	0	1
FALSE	21334	0	0	0	0	5
FALSE	21339	0	0	0	0	1
FALSE	21341	0	549	0	0	2
FALSE	21347	0	14	0	1	3
FALSE	21348	0	1	0	0	69
FALSE	21350	0	0	0	0	12
FALSE	21354	0	100	0	0	0
FALSE	21355	0	0	0	0	78
FALSE	21356	0	0	0	0	2
FALSE	21360	0	0	0	0	3
FALSE	21368	0	0	0	0	108
FALSE	21370	0	13	0	1	19
FALSE	21371	0	0	0	1	313
FALSE	21375	0	0	0	0	39
FALSE	21376	0	0	0	0	9
FALSE	21377	0	0	0	0	9
FALSE	21380	0	0	0	0	4
FALSE	21384	0	0	0	0	30
FALSE	21387	0	0	0	0	6

FALSE	21388	0	0	0	0	184
FALSE	21390	0	412	0	1	1
FALSE	21394	0	0	0	0	7
FALSE	21395	0	180	0	0	0
FALSE	21397	0	0	0	0	16
FALSE	21405	0	0	0	0	61
FALSE	21420	0	345	0	1	0
FALSE	21422	0	7	0	0	0
FALSE	21423	0	0	0	0	14
FALSE	21438	0	152	0	0	0
FALSE	21439	0	190	0	0	0
FALSE	21440	0	62	0	0	0
FALSE	21441	0	2	0	0	1
FALSE	21444	0	8	0	0	0
FALSE	21446	0	0	0	0	2
FALSE	21710	0	0	0	1	0
FALSE	21720	0	270	0	1	158
FALSE	21725	0	0	0	0	228
FALSE	21727	0	28	0	0	0
FALSE	21731	0	141	0	0	0
FALSE	21735	0	0	0	1	320
FALSE	21740	0	495	0	1	248
FALSE	21753	0	0	0	0	24
FALSE	21800	0	12	0	0	0
FALSE	21805	0	14	0	0	0
FALSE	21810	0	8	0	0	0
FALSE	21811	0	1	0	0	0
FALSE	21900	0	0	0	0	2
FALSE	21921	0	17	0	0	0
FALSE	21932	0	0	0	0	127
FALSE	21935	0	0	0	1	0
FALSE	22010	0	5	0	0	0
FALSE	22175	0	234	0	0	0
FALSE	22178	0	0	0	0	5
FALSE	22200	0	155	0	0	109
FALSE	22201	0	0	0	0	25
FALSE	22205	0	0	0	0	24
FALSE	22206	0	7	0	0	0
FALSE	22213	0	0	0	0	9
FALSE	22214	0	7	0	0	0
FALSE	22215	0	339	0	0	0
FALSE	22216	0	178	0	0	0
FALSE	22217	0	2	0	0	0
FALSE	22219	0	24	0	0	0
FALSE	22220	0	349	0	0	0
FALSE	22222	0	6	0	0	0
FALSE	22223	0	33	0	0	0
FALSE	22224	0	1	0	0	0
FALSE	22226	0	1	0	0	1
FALSE	22228	0	198	0	0	0
FALSE	22229	0	288	0	0	0
FALSE	22231	0	34	0	0	0
FALSE	22232	0	2	0	0	0
FALSE	22233	0	22	0	0	0

```
FALSE        22234   0       8   0   0   0
FALSE        22235   0     162   0   0   0
FALSE        22236   0     181   0   0   1
FALSE        22239   0       1   0   0   0
FALSE        22240   0       1   0   0   0
FALSE        22241   0      31   0   0   0
FALSE        22244   0      24   0   0   0
FALSE        22245   0      10   0   0   0
FALSE        22246   0       2   0   0   0
FALSE        22249   0      32   0   0   0
FALSE        22250   0      26   0   0   0
FALSE        22252   0       2   0   0   0
FALSE        22253   0       7   0   0   0
FALSE        22255   0       2   0   0   0
FALSE        22260   0      26   0   0   0
FALSE        22261   0       6   0   0   0
FALSE        22262   0      49   0   0   0
FALSE        22265   0       0   0   0   1
FALSE        22271   0       5   0   0   0
FALSE        22276   0      13   0   0   0
FALSE        22277   0       2   0   0   0
FALSE        22281   0       3   0   0   0
FALSE        22282   0       7   0   0   0
FALSE        22283   0       2   0   0   0
FALSE        22300   0       6   0   0   0
FALSE        22310   0      33   0   0   0
FALSE        22320   0      34   0   0   0
FALSE        22390   0       1   0   0   0
FALSE        22410   0       7   0   0   0
FALSE        22420   0      24   0   0   0
FALSE        22461   0       1   0   0   0
FALSE  [ reached getOption("max.print") -- omitted 750 rows ]
```

Look through new and old 2021 data and note where differences are

Please let me know if there are any differences between these datasets that are unacceptable (e.g., should be more like the old data or completely changed) and what tables and columns I should reference to fix these issues.

Note that this will only print where these trouble makers occurred if they occurred.

- characters: x = new; y = old; new | old
- numbers: a summary table of where these differences occurred with additional columns for context (common_name, species_code, year, and srvy)

Things to look out for in scientific names (that we might want to make formative changes in "RACEBASE.SPECIES_CLASSIFICATION$report_name_scientific"):

- NOTE I already removed commas (see note above; bet it doesn't matter) from old data
- "unid." and "uid" in new data vs [nothing] in old data
- where "egg"/"egg case" is included in scientific names of old data and if it should be included in the new data. Should these "egg"/"egg case" notes only be included in the common name?

- ownership dispute between McLean and Clark (e.g., “Neptunea sp. E (McLean and Clark)” and “Neptunea sp. E (Clark and McLean)” and . . . “(McLean & Clark))

Other notes

- -9999 –> NA from cpue_noha and cpue_kgha in the old data so we can compare new and old more easily

```
col <- gsub(pattern = ".x", replacement = "",
            x = names(dat)[grep(pattern = ".x", x = names(dat), fixed = TRUE)])
col <- col[-length(col)]
# # Compare x to y
for(i in 1:length(col)) {
  a <- diff_between_cols(dat = dat, colx = paste0(col[i], ".x"), coly = paste0(col[i], ".y"))
  if (class(a) == "table") {
     print(paste0("----- Differences between '", col[i], ".x' (new) and '", col[i], ".y' (old) -----"))
     print(diff_between_cols(dat = dat, colx = paste0(col[i], ".x"), coly = paste0(col[i], ".y")))
  }
}
```

```
FALSE [1] "----- Differences between 'survey.x' (new) and 'survey.y' (old) -----"
FALSE
FALSE                                                    Aleutian Islands Bottom Trawl Survey | Aleuti
FALSE
FALSE                              Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey | Southeastern
FALSE
FALSE                                                       Gulf of Alaska Bottom Trawl Survey | Gul
FALSE
FALSE Northern Bering Sea Crab/Groundfish Survey - Eastern Bering Sea Shelf Survey Extension | Northern
FALSE
FALSE [1] "----- Differences between 'common_name.x' (new) and 'common_name.y' (old) -----"
FALSE
FALSE          Bering scallop | Iceland scallop     chrysaora jelly | chrysaora jellyfish
FALSE                                      77                                        873
FALSE    gray mushroom coral | gray Anthomastus          lion's mane jelly | lion's mane
FALSE                                      21                                       1906
FALSE orange finger sponge | soft finger sponge               Pacific capelin | capelin
FALSE                                     235                                       5934
FALSE    rams-horn snail | rams-horn hairysnail      red mushroom coral | red anthomastus
FALSE                                      15                                        129
FALSE [1] "----- Differences between 'scientific_name.x' (new) and 'scientific_name.y' (old) -----"
FALSE
FALSE          Albatrossia pectoralis | Coryphaenoides pectoralis
FALSE                                                      1116
FALSE                   Aspidophoroides olrikii | Ulcina olrikii
FALSE                                                        72
FALSE                    Beringius behringi | Beringius beringii
FALSE                                                       992
FALSE              Glebocarcinus oregonensis | Cancer oregonensis
FALSE                                                       991
FALSE      Heteropolypus (=Anthomastus) sp. A | Anthomastus sp. A
FALSE                                                       129
FALSE      Heteropolypus (=Anthomastus) sp. B | Anthomastus sp. B
```

```
FALSE                                                    21
FALSE       Hippasteria lepidonotus | Cryptopeltaster lepidonotus
FALSE                                                     3
FALSE                    Lebbeus grandimanus | Lebbeus grandimana
FALSE                                                     1
FALSE                  Liparis bristolensis | Liparis bristolense
FALSE                                                     1
FALSE                           Lunatia lewisii | Euspira lewisii
FALSE                                                    11
FALSE        Mallotus catervarius (=villosus) | Mallotus villosus
FALSE                                                  5934
FALSE                     Metacarcinus gracilis | Cancer gracilis
FALSE                                                     1
FALSE                     Metacarcinus magister | Cancer magister
FALSE                                                    97
FALSE      Neptunea alexeyevi | Neptunea sp. F (Clark and McLean)
FALSE                                                     1
FALSE Onchidiopsis clarki | Onchidiopsis sp. B (Clark and McLean)
FALSE                                                    15
FALSE                  Otukaia beringensis | Otukaia kiheiziebisu
FALSE                                                     2
FALSE              Psychrolutes sigalutes | Gilbertidia sigalutes
FALSE                                                     2
FALSE                          Romaleon branneri | Cancer branneri
FALSE                                                     5
FALSE                 Torellivelutina ammonia | Torellia ammonia
FALSE                                                    15
FALSE                Tripoplax abyssicola | Lepidozona abyssicola
FALSE                                                     6
FALSE                Tripoplax beringiana | Lepidozona beringiana
FALSE                                                     6
FALSE                      Tripoplax trifida | Lepidozona trifida
FALSE                                                    19
FALSE [1] "----- Differences between 'cpue_kgha.x' (new) and 'cpue_kgha.y' (old) -----"
FALSE              common_name.x     species_code         year       srvy          station             s
FALSE  walleye pollock    : 26679   21740  : 26679   Min.   :1982   AI :140362   Length:870605      Min
FALSE  Pacific cod        : 25762   21720  : 25762   1st Qu.:1996   EBS:455440   Class :character   1st
FALSE  arrowtooth flounder: 22078   10110  : 22078   Median :2005   GOA:250332   Mode  :character   Medi
FALSE  Pacific halibut    : 21990   10120  : 21990   Mean   :2004   NBS: 24471                      Mean
FALSE  flathead sole      : 19239   10130  : 19239   3rd Qu.:2013                                   3rd
FALSE  (Other)            :661303   10261  : 13872   Max.   :2021                                   Max
FALSE  NA's               : 93554   (Other):740985
FALSE      cruise        cpue_kgha.x        cpue_kgha.y         difference
FALSE  Min.   :198201   Min.   :    0.00   Min.   :    0.00   Min.   :-0.002258
FALSE  1st Qu.:199601   1st Qu.:    0.05   1st Qu.:    0.05   1st Qu.:-0.000025
FALSE  Median :200501   Median :    0.44   Median :    0.44   Median :-0.000001
FALSE  Mean   :200387   Mean   :   11.80   Mean   :   11.80   Mean   : 0.000025
FALSE  3rd Qu.:201301   3rd Qu.:    3.41   3rd Qu.:    3.41   3rd Qu.: 0.000025
FALSE  Max.   :202102   Max.   :32262.35   Max.   :32262.35   Max.   :18.865709
FALSE
FALSE [1] "----- Differences between 'cpue_noha.x' (new) and 'cpue_noha.y' (old) -----"
FALSE              common_name.x     species_code         year       srvy          station             s
FALSE  walleye pollock    : 26656   21740  : 26656   Min.   :1982   AI : 61441   Length:408330      Min
FALSE  Pacific cod        : 25718   21720  : 25718   1st Qu.:1994   EBS:193269   Class :character   1st
```

```
FALSE  arrowtooth flounder: 22033   10110  : 22033   Median :2004   GOA:143645   Mode  :character   Medi
FALSE  Pacific halibut    : 21986   10120  : 21986   Mean   :2003   NBS:  9975                      Mean
FALSE  flathead sole      : 19222   10130  : 19222   3rd Qu.:2012                                  3rd
FALSE  (Other)            :287848   10261  : 13864   Max.   :2021                                  Max
FALSE  NA's               :  4867   (Other):278851
FALSE      cruise        cpue_noha.x        cpue_noha.y         difference
FALSE  Min.   :198201   Min.   :    0.04   Min.   :    0.04   Min.   :-0.000050
FALSE  1st Qu.:199401   1st Qu.:    0.63   1st Qu.:    0.63   1st Qu.:-0.000025
FALSE  Median :200401   Median :    2.45   Median :    2.45   Median : 0.000001
FALSE  Mean   :200290   Mean   :   44.22   Mean   :   44.22   Mean   : 0.000037
FALSE  3rd Qu.:201201   3rd Qu.:   13.31   3rd Qu.:   13.31   3rd Qu.: 0.000025
FALSE  Max.   :202102   Max.   :44817.02   Max.   :44817.02   Max.   :15.062311
FALSE
```

```
dat0 <- dat %>%
  dplyr::mutate(
    dplyr::across(dplyr::starts_with("cpue_"), round, digits = 2),
    cpue_kgha_diff = cpue_kgha.x-cpue_kgha.y,
    cpue_noha_diff = cpue_noha.x-cpue_noha.y) %>%
  dplyr::select(year, srvy, cruise, stratum, station, haul, vessel_id,
                # latitude_dd.x, longitude_dd.x,
                species_code, common_name.x, scientific_name.x,
                cpue_kgha.x, cpue_kgha.y, cpue_kgha_diff,
                cpue_noha.x, cpue_noha.y, cpue_noha_diff) %>%
  dplyr::filter(cpue_kgha_diff >= 0.01 | cpue_noha_diff >= 0.01) %>%
  dplyr::arrange(-cpue_kgha_diff, -cpue_noha_diff)

# dat0 <- dat0[,order(colnames(dat0))]

readr::write_csv(dat0,
  file = here::here("output", Sys.Date(), "diff_xnew_yold.csv"))

summary(dat0)
```

```
FALSE       year          srvy               cruise          stratum        station               haul
FALSE  Min.   :1982   Length:2139        Min.   :198201   Min.   : 10.0   Length:2139        Min.   :
FALSE  1st Qu.:1998   Class :character   1st Qu.:199851   1st Qu.: 31.0   Class :character   1st Qu.: 6
FALSE  Median :2006   Mode  :character   Median :200601   Median : 61.0   Mode  :character   Median :11
FALSE  Mean   :2005                      Mean   :200503   Mean   :149.5                      Mean   :12
FALSE  3rd Qu.:2013                      3rd Qu.:201301   3rd Qu.:214.0                      3rd Qu.:17
FALSE  Max.   :2021                      Max.   :202102   Max.   :794.0                      Max.   :35
FALSE
FALSE    vessel_id      species_code   common_name.x      scientific_name.x   cpue_kgha.x        cpue_kg
FALSE  Min.   :  1.0   Min.   :   21   Length:2139        Length:2139        Min.   :   0.000   Min.
FALSE  1st Qu.: 88.0   1st Qu.:20061   Class :character   Class :character   1st Qu.:   0.020   1st Qu.
FALSE  Median : 94.0   Median :30020   Mode  :character   Mode  :character   Median :   0.260   Median
FALSE  Mean   :118.7   Mean   :42568                                         Mean   :   6.678   Mean
FALSE  3rd Qu.:147.0   3rd Qu.:71310                                         3rd Qu.:   1.485   3rd Qu.
FALSE  Max.   :560.0   Max.   :99998                                         Max.   :3190.090   Max.
FALSE
FALSE  cpue_kgha_diff      cpue_noha.x        cpue_noha.y       cpue_noha_diff
FALSE  Min.   : 0.00000   Min.   :   0.080   Min.   :   0.080   Min.   :-0.0100
FALSE  1st Qu.: 0.00000   1st Qu.:   0.430   1st Qu.:   0.430   1st Qu.: 0.0000
FALSE  Median : 0.01000   Median :   1.080   Median :   1.100   Median : 0.0000
```

```
FALSE  Mean   : 0.01765   Mean   :  18.618   Mean   :  21.961   Mean   : 0.0178
FALSE  3rd Qu.: 0.01000   3rd Qu.:   3.945   3rd Qu.:   4.093   3rd Qu.: 0.0100
FALSE  Max.   :18.87000   Max.   :4776.960   Max.   :4776.960   Max.   :15.0600
FALSE                     NA's   :213        NA's   :975        NA's   :975
```

How do CPUE values compare with our design-based CPUE estimates?

Here I am compiling the EBS, NBS, GOA, AI verified, design-based datasets

```
design_based1 <- dplyr::bind_rows(
  readr::read_csv(file = "../data/oracle/cpue_ai.csv"),
  readr::read_csv(file = "../data/oracle/cpue_goa.csv")) %>%
  janitor::clean_names() %>%
    dplyr::left_join(
      x = .,
      y = haul %>%
        dplyr::select(hauljoin, stationid, start_latitude, start_longitude),
      by = "hauljoin")  %>%
    dplyr::left_join(
      x = .,
      y = spp_info %>%
        dplyr::select(common_name, scientific_name, species_code),
      by = "species_code") %>%
  dplyr::select(-x1, -catchjoin) %>%
  dplyr::rename(latitude = start_latitude,
                longitude = start_longitude,
                weight_kg = weight,
                count = number_fish) %>%
  dplyr::mutate(cpue_kgha = wgtcpue/100, # denominator orig in km2
                cpue_noha = numcpue/100, # denominator orig in km2
                area_swept_ha = effort*100) # denominator orig in km2

design_based2 <- dplyr::bind_rows(
    readr::read_csv(file = "../data/oracle/cpue_ebs_plusnw.csv") %>%
      dplyr::mutate(survey = "EBS"),
    readr::read_csv(file = "../data/oracle/cpue_nbs.csv") %>%
      dplyr::mutate(survey = "NBS")) %>%
  janitor::clean_names() %>%
  dplyr::select(-x1)  %>%
  dplyr::left_join(
      x = .,
      y = haul %>%
        dplyr::select(hauljoin, cruise, distance_fished, net_width),
      by = "hauljoin") %>%
  dplyr::mutate(weight_kg = NA,
                count = NA) %>%
  dplyr::rename(area_swept_ha = area_fished_ha,
                scientific_name = species_name)

design_based <- dplyr::bind_rows(design_based1, design_based2) %>%
  dplyr::rename(srvy = survey,
```

```
                vessel_id = vessel,
                distance_fished_km = distance_fished,
                net_width_m = net_width,
                station = stationid,
                latitude_dd = latitude,
                longitude_dd = longitude)
dim(design_based) # before summary
```

```
FALSE [1] 3113503      23
```

```
# summarising bc there are duplicates in some of the AI and GOA(?) design based indicies
design_based <- design_based %>%
  dplyr::group_by(srvy, year, hauljoin, vessel_id, cruise, haul, stratum, station,
                  common_name, scientific_name, species_code) %>%
  dplyr::summarise(distance_fished_km = mean(distance_fished_km, na.rm = TRUE),
                   net_width_m = mean(net_width_m, na.rm = TRUE),
                   weight_kg = sum(weight_kg, na.rm = TRUE),
                   count = sum(count, na.rm = TRUE),
                  latitude_dd = mean(latitude_dd, na.rm = TRUE),
                  longitude_dd = mean(longitude_dd, na.rm = TRUE),
                  cpue_kgha = sum(cpue_kgha, na.rm = TRUE),
                  cpue_noha = sum(cpue_noha, na.rm = TRUE),
                  area_swept_ha = sum(area_swept_ha, na.rm = TRUE))

dim(design_based) # after summary
```

```
FALSE [1] 3113502      20
```

```
design_based
```

```
FALSE # A tibble: 3,113,502 x 20
FALSE # Groups:   srvy, year, hauljoin, vessel_id, cruise, haul, stratum, station, common_name, scienti
FALSE #   [3,113,502]
FALSE    srvy   year hauljoin vessel_id cruise  haul stratum station common_name          scientific_nar
FALSE    <chr> <dbl>    <dbl>     <dbl>  <dbl> <dbl>   <dbl> <chr>   <chr>                <chr>
FALSE  1 AI     1980    27698        31 198002    42     622 <NA>    Alaska skate         Bathyraja parr
FALSE  2 AI     1980    27698        31 198002    42     622 <NA>    Aleutian skate       Bathyraja aleu
FALSE  3 AI     1980    27698        31 198002    42     622 <NA>    antlered sculpin     Enophrys dicer
FALSE  4 AI     1980    27698        31 198002    42     622 <NA>    armorhead sculpin    Gymnocanthus g
FALSE  5 AI     1980    27698        31 198002    42     622 <NA>    arrowtooth flounder  Atheresthes st
FALSE  6 AI     1980    27698        31 198002    42     622 <NA>    Atka mackerel        Pleurogrammus
FALSE  7 AI     1980    27698        31 198002    42     622 <NA>    Bering skate         Bathyraja int
FALSE  8 AI     1980    27698        31 198002    42     622 <NA>    big skate            Beringraja bi
FALSE  9 AI     1980    27698        31 198002    42     622 <NA>    bigmouth sculpin     Hemitripterus
FALSE 10 AI     1980    27698        31 198002    42     622 <NA>    blackspotted rockfi~ Sebastes melar
FALSE # ... with 3,113,492 more rows, and 9 more variables: distance_fished_km <dbl>, net_width_m <dbl>
FALSE #   weight_kg <dbl>, count <dbl>, latitude_dd <dbl>, longitude_dd <dbl>, cpue_kgha <dbl>, cpue_nol
FALSE #   area_swept_ha <dbl>
```

And now I am joining the new data and the designed based data together to see where differences arise.

```r
dat <- dplyr::full_join(
  x = data_new %>%
    dplyr::select(common_name, species_code, scientific_name,
                  cpue_kgha, cpue_noha,
                  area_swept_ha, distance_fished_km, net_width_m,
                  haul, cruise, station, stratum, srvy,
                  count, weight_kg,
                  vessel_id, latitude_dd, longitude_dd, year),
  y = design_based %>%
    dplyr::select(common_name, species_code, scientific_name,
                  cpue_kgha, cpue_noha,
                  area_swept_ha, distance_fished_km, net_width_m,
                  haul, cruise, station, stratum, srvy,
                  count, weight_kg,
                  vessel_id, latitude_dd, longitude_dd, year),
  by = c("year", "haul", "cruise", "srvy", "species_code", "vessel_id", "stratum", "station")) %>%
  dplyr::mutate(dplyr::across(dplyr::starts_with("cpue_"), round, digits = 2),
                dplyr::across(dplyr::starts_with("area_swept_"), round, digits = 4))

col <- gsub(pattern = ".x", replacement = "",
            x = names(dat)[grep(pattern = ".x", x = names(dat), fixed = TRUE)])

## Compare x to y
for(i in 1:length(col)) {
  a <- diff_between_cols(dat = dat,
                         colx = paste0(col[i], ".x"),
                         coly = paste0(col[i], ".y"))
  if (class(a) == "table") {
     print(paste0("----- Differences between '", col[i], ".x' (public) and '", col[i], ".y' (design) ---
     print(diff_between_cols(dat = dat, colx = paste0(col[i], ".x"), coly = paste0(col[i], ".y")))
  }
}
```

```
FALSE [1] "----- Differences between 'cpue_kgha.x' (public) and 'cpue_kgha.y' (design) -----"
FALSE          common_name.x  species_code      year       srvy       station             stratum
FALSE  Pacific cod    : 76   21720  : 76   Min.   :1982   AI : 24   Length:886         Min.   : 10.00
FALSE  Pacific halibut: 60   10120  : 60   1st Qu.:1994   EBS:712   Class :character   1st Qu.: 31.00
FALSE  flathead sole  : 58   10130  : 58   Median :2004   GOA: 69   Mode  :character   Median : 42.00
FALSE  walleye pollock: 52   21740  : 52   Mean   :2003   NBS: 81                      Mean   : 69.24
FALSE  yellowfin sole : 51   10210  : 51   3rd Qu.:2013                                3rd Qu.: 61.00
FALSE  (Other)        :575   471    : 43   Max.   :2021                                Max.   :793.00
FALSE  NA's           : 14   (Other):546
FALSE      cruise        cpue_kgha.x        cpue_kgha.y         difference
FALSE  Min.   :198201   Min.   :   0.000   Min.   :   0.000   Min.   :-7.55000
FALSE  1st Qu.:199401   1st Qu.:   0.242   1st Qu.:   0.285   1st Qu.:-0.01000
FALSE  Median :200401   Median :   2.020   Median :   2.100   Median :-0.01000
FALSE  Mean   :200309   Mean   :  24.497   Mean   :  24.548   Mean   :-0.05058
FALSE  3rd Qu.:201301   3rd Qu.:  11.065   3rd Qu.:  11.060   3rd Qu.: 0.01000
FALSE  Max.   :202102   Max.   :6363.010   Max.   :6363.020   Max.   : 0.46000
FALSE
FALSE [1] "----- Differences between 'cpue_noha.x' (public) and 'cpue_noha.y' (design) -----"
FALSE          common_name.x  species_code      year       srvy       station             stratum
FALSE  walleye pollock: 73   21740  : 73   Min.   :1982   AI : 24   Length:829         Min.   : 10.00
```

```
FALSE  Pacific cod    : 55   10120  : 55   1st Qu.:1994   EBS:665   Class :character   1st Qu.: 31.00
FALSE  Pacific halibut: 55   21720  : 55   Median :2004   GOA: 61   Mode  :character   Median : 43.00
FALSE  flathead sole  : 54   10130  : 54   Mean   :2003   NBS: 79                      Mean   : 70.75
FALSE  Alaska skate   : 49   471    : 49   3rd Qu.:2013                                3rd Qu.: 62.00
FALSE  (Other)        :523   10285  : 46   Max.   :2021                                Max.   :721.00
FALSE  NA's           : 20   (Other):497
FALSE      cruise        cpue_noha.x       cpue_noha.y        difference
FALSE  Min.   :198201   Min.   :   0.08   Min.   :   0.00   Min.   :-1122.410
FALSE  1st Qu.:199401   1st Qu.:   0.81   1st Qu.:   0.96   1st Qu.:   -0.010
FALSE  Median :200401   Median :   3.19   Median :   3.34   Median :   -0.010
FALSE  Mean   :200336   Mean   :  56.64   Mean   :  58.33   Mean   :   -1.688
FALSE  3rd Qu.:201301   3rd Qu.:  24.62   3rd Qu.:  25.25   3rd Qu.:    0.010
FALSE  Max.   :202102   Max.   :3815.46   Max.   :3815.45   Max.   :    6.410
FALSE
FALSE [1] "----- Differences between 'area_swept_ha.x' (public) and 'area_swept_ha.y' (design) -----"
FALSE              common_name.x  species_code      year       srvy       station             stratum
FALSE  walleye pollock    : 39   21740  : 39   Min.   :1982   AI : 53   Length:525         Min.   : 10.0
FALSE  Pacific cod        : 34   21720  : 34   1st Qu.:1993   EBS:167   Class :character   1st Qu.: 41.0
FALSE  Pacific halibut    : 31   10120  : 31   Median :2003   GOA:261   Mode  :character   Median : 90.0
FALSE  arrowtooth flounder: 29   10110  : 29   Mean   :2003   NBS: 44                      Mean   :108.1
FALSE  flathead sole      : 26   10130  : 26   3rd Qu.:2015                                3rd Qu.:140.0
FALSE  (Other)            :355   10200  : 20   Max.   :2021                                Max.   :330.0
FALSE  NA's               : 11   (Other):346
FALSE      cruise       area_swept_ha.x  area_swept_ha.y    difference
FALSE  Min.   :198201   Min.   : 2.094   Min.   : 2.094   Min.   :-9.09560
FALSE  1st Qu.:199301   1st Qu.: 2.413   1st Qu.: 2.413   1st Qu.:-0.00010
FALSE  Median :200301   Median : 2.884   Median : 2.884   Median : 0.00010
FALSE  Mean   :200334   Mean   : 3.647   Mean   : 3.665   Mean   :-0.01729
FALSE  3rd Qu.:201501   3rd Qu.: 4.514   3rd Qu.: 4.514   3rd Qu.: 0.00010
FALSE  Max.   :202101   Max.   :12.313   Max.   :18.191   Max.   : 0.00010
FALSE
FALSE [1] "----- Differences between 'count.x' (public) and 'count.y' (design) -----"
FALSE           common_name.x     species_code        year       srvy          station             stratu
FALSE  walleye pollock: 14510   21740  :14510   Min.   :1982   AI :    20   Length:165075      Min.   :
FALSE  Pacific cod    : 14271   21720  :14271   1st Qu.:1994   EBS:147730   Class :character   1st Qu.:
FALSE  flathead sole  : 11091   10130  :11091   Median :2005   GOA:    63   Mode  :character   Median :
FALSE  Alaska plaice  : 10392   10285  :10392   Mean   :2004   NBS: 17262                      Mean   :
FALSE  yellowfin sole : 10316   10210  :10316   3rd Qu.:2014                                   3rd Qu.:
FALSE  (Other)        :101414   10120  :10039   Max.   :2021                                   Max.   :
FALSE  NA's           :  3081   (Other):94456
FALSE      cruise          count.x           count.y            difference
FALSE  Min.   :198201   Min.   :    0.0   Min.   :   0.0000   Min.   :-1984.0
FALSE  1st Qu.:199401   1st Qu.:    4.0   1st Qu.:   0.0000   1st Qu.:    4.0
FALSE  Median :200501   Median :   17.0   Median :   0.0000   Median :   17.0
FALSE  Mean   :200377   Mean   :  267.9   Mean   :   0.0196   Mean   :  267.9
FALSE  3rd Qu.:201401   3rd Qu.:  111.0   3rd Qu.:   0.0000   3rd Qu.:  111.0
FALSE  Max.   :202102   Max.   :68196.0   Max.   :2017.0000   Max.   :68196.0
FALSE
FALSE [1] "----- Differences between 'weight_kg.x' (public) and 'weight_kg.y' (design) -----"
FALSE           common_name.x     species_code        year       srvy          station             stratu
FALSE  walleye pollock: 14517   21740  :14517   Min.   :1982   AI :    27   Length:168115      Min.   :
FALSE  Pacific cod    : 14278   21720  :14278   1st Qu.:1995   EBS:147842   Class :character   1st Qu.:
FALSE  flathead sole  : 11098   10130  :11098   Median :2005   GOA:    67   Mode  :character   Median :
FALSE  Alaska plaice  : 10400   10285  :10400   Mean   :2004   NBS: 20179                      Mean   :
```

```
FALSE  yellowfin sole : 10324   10210  :10324   3rd Qu.:2014                                    3rd Qu.:
FALSE  (Other)        :104068   10120  :10040   Max.   :2021                                    Max.   :
FALSE  NA's           :  3430   (Other):97458
FALSE      cruise        weight_kg.x         weight_kg.y          difference
FALSE  Min.   :198201   Min.   :    0.001   Min.   :    0.000   Min.   : -21.228
FALSE  1st Qu.:199501   1st Qu.:    2.268   1st Qu.:    0.000   1st Qu.:   2.266
FALSE  Median :200501   Median :   13.126   Median :    0.000   Median :  13.110
FALSE  Mean   :200398   Mean   :   96.427   Mean   :    0.725   Mean   :  95.702
FALSE  3rd Qu.:201401   3rd Qu.:   59.928   3rd Qu.:    0.000   3rd Qu.:  59.920
FALSE  Max.   :202102   Max.   :26809.310   Max.   :26809.308   Max.   :11465.229
FALSE
```

```
dat0 <- dat %>%
  dplyr::mutate(
    cpue_kgha_diff = cpue_kgha.x-cpue_kgha.y,
    cpue_noha_diff = cpue_noha.x-cpue_noha.y,
    area_swept_ha_diff = area_swept_ha.x-area_swept_ha.y) %>%
  dplyr::select(-distance_fished_km.x, -net_width_m.x, -latitude_dd.x, -longitude_dd.x,
                -distance_fished_km.y, -net_width_m.y, -latitude_dd.y, -longitude_dd.y,
                -common_name.y, -scientific_name.y) %>%
  dplyr::filter(cpue_kgha_diff >= 0.01 |
                  # ((is.na(cpue_kgha_diff) & !is.na(cpue_kgha.x)) |
                  #    (is.na(cpue_kgha_diff) & !is.na(cpue_kgha.y))) |
                  cpue_noha_diff >= 0.01 | #is.na(cpue_noha_diff) |
                  # ((is.na(cpue_noha_diff) & !is.na(cpue_noha.x)) |
                  #    (is.na(cpue_noha_diff) & !is.na(cpue_noha.y))) |
                  area_swept_ha_diff >= 0.01 # |
                  # ((is.na(area_swept_ha_diff) & !is.na(area_swept_ha.x)) |
                  #    (is.na(area_swept_ha_diff) & !is.na(area_swept_ha.y)))
                ) %>%
  dplyr::arrange(-cpue_kgha_diff, -cpue_noha_diff, -area_swept_ha_diff) %>%
  dplyr::relocate(year, srvy, cruise, haul, station, stratum, vessel_id,
                  species_code, common_name.x,
                  area_swept_ha.x, area_swept_ha.y, area_swept_ha_diff,
                  weight_kg.x, weight_kg.y, cpue_kgha.x, cpue_kgha.y, cpue_kgha_diff,
                  count.x, count.y, cpue_kgha.x, cpue_kgha.y, cpue_kgha_diff)

# dat0 <- dat0[,order(colnames(dat0))]

readr::write_csv(dat0,
  file = here::here("output", Sys.Date(), "diff_xnew_ydesign.csv"))

summary(dat0)
```

```
FALSE       year          srvy               cruise           haul          station             stratu
FALSE  Min.   :1982   Length:462         Min.   :198203   Min.   :  1.00   Length:462         Min.   :
FALSE  1st Qu.:1997   Class :character   1st Qu.:199726   1st Qu.: 48.25   Class :character   1st Qu.:
FALSE  Median :2008   Mode  :character   Median :200751   Median : 95.00   Mode  :character   Median :
FALSE  Mean   :2005                      Mean   :200549   Mean   :100.42                      Mean   :
FALSE  3rd Qu.:2015                      3rd Qu.:201501   3rd Qu.:151.00                      3rd Qu.:
FALSE  Max.   :2021                      Max.   :202102   Max.   :285.00                      Max.   :7
FALSE
FALSE    vessel_id      species_code   common_name.x      area_swept_ha.x  area_swept_ha.y  area_swept_
FALSE  Min.   :  1.0   Min.   :  420   Length:462         Min.   : 1.839   Min.   : 1.839   Min.   :0
```

```
FALSE  1st Qu.: 88.0   1st Qu.:10120   Class :character   1st Qu.: 4.416   1st Qu.: 4.416   1st Qu.:0
FALSE  Median : 89.0   Median :10261   Mode  :character   Median : 4.739   Median : 4.739   Median :0
FALSE  Mean   :101.3   Mean   :18863                      Mean   : 4.686   Mean   : 4.686   Mean   :0
FALSE  3rd Qu.:134.0   3rd Qu.:21438                      3rd Qu.: 5.095   3rd Qu.: 5.095   3rd Qu.:0
FALSE  Max.   :554.0   Max.   :95036                      Max.   :16.018   Max.   :16.018   Max.   :0
FALSE
FALSE   weight_kg.x        weight_kg.y       cpue_kgha.x        cpue_kgha.y       cpue_kgha_diff
FALSE  Min.   :   0.005   Min.   :   0.00   Min.   :   0.000   Min.   :   0.000   Min.   :0.000000
FALSE  1st Qu.:   1.052   1st Qu.:   0.00   1st Qu.:   0.230   1st Qu.:   0.212   1st Qu.:0.000000
FALSE  Median :   6.020   Median :   0.00   Median :   1.310   Median :   1.305   Median :0.010000
FALSE  Mean   :  69.565   Mean   :  16.16   Mean   :  18.310   Mean   :  18.303   Mean   :0.007273
FALSE  3rd Qu.:  30.261   3rd Qu.:   0.00   3rd Qu.:   6.805   3rd Qu.:   6.805   3rd Qu.:0.010000
FALSE  Max.   :7239.490   Max.   :7239.49   Max.   :3226.860   Max.   :3226.860   Max.   :0.460000
FALSE
FALSE     count.x           count.y        scientific_name.x   cpue_noha.x          hauljoin        cpu
FALSE  Min.   :   0.00   Min.   :   0.00   Length:462         Min.   :   0.120   Min.   : -20809   Min.
FALSE  1st Qu.:   3.00   1st Qu.:   0.00   Class :character   1st Qu.:   0.630   1st Qu.: -14564   1st
FALSE  Median :   7.00   Median :   0.00   Mode  :character   Median :   1.510   Median :  -5412   Medi
FALSE  Mean   : 171.39   Mean   :  18.93                      Mean   :  41.244   Mean   : 249346   Mean
FALSE  3rd Qu.:  38.75   3rd Qu.:   0.00                      3rd Qu.:   8.365   3rd Qu.: 771798   3rd
FALSE  Max.   :8560.00   Max.   :8560.00                      Max.   :3815.460   Max.   :1195900   Max.
FALSE                                                         NA's   :2
FALSE  cpue_noha_diff
FALSE  Min.   :0.00000
FALSE  1st Qu.:0.00000
FALSE  Median :0.01000
FALSE  Mean   :0.02835
FALSE  3rd Qu.:0.01000
FALSE  Max.   :6.41000
FALSE  NA's   :2
```

What does the new 2021 data look like?

Look through new 2021 data and see if there were any -9999, NA, " ", or 0's

Note that this will only print where these trouble makers occurred if they occurred.

```
trouble <- c(-9999, NA, "", 0)
cols <- names(data_new)
cols <- cols[cols != "date"]

for (i in 1:length(trouble)){
  print(paste0("------------- Are there any '", trouble[i], "' in these columns? -------------"))
  for (ii in 1:length(cols)) {
     a <- wheres_trouble(dat = data_new, col = cols[ii], trouble = trouble[i])
      if (class(a) == "list") {
        print(a)
      }
  }
}
```

```
FALSE [1] "------------- Are there any '-9999' in these columns? -------------"
```

```
FALSE $‘Yes, ’-9999’ occured 43 times in the column ’hauljoin’.‘
FALSE       srvy
FALSE year   EBS
FALSE   2013  43
FALSE
FALSE [1] "------------- Are there any ’NA’ in these columns? -------------"
FALSE $‘Yes, ’NA’ occured 3539 times in the column ’taxon_confidence’.‘
FALSE       srvy
FALSE year    AI BSSlope EBS GOA NBS
FALSE   1983   0       0   2   0   0
FALSE   1984   0       0   0  10   0
FALSE   1985   0       0   2   0   0
FALSE   1986   1       0   0   0   0
FALSE   1987   0       0   1   3   0
FALSE   1990   0       0   1   2   0
FALSE   1991   1       0   0   0   0
FALSE   1992   0       0   1   0   0
FALSE   1994   2       0   2   0   0
FALSE   1996   0       0   0   2   0
FALSE   1997  11       0   2   0   0
FALSE   1999   0       0   2   5   0
FALSE   2000  11       0   1   0   0
FALSE   2001   0       0   0   3   0
FALSE   2002  12       1   2   0   0
FALSE   2003   0       0   3  10   0
FALSE   2004  24      10   7   0   0
FALSE   2005   0       0   7  23   0
FALSE   2006  24       0   1   0   0
FALSE   2007   0       0   3  25   0
FALSE   2008   0       5  30   0   0
FALSE   2009   0       0  52  12   0
FALSE   2010  53      11 102   0 132
FALSE   2011   0       0  36   9   0
FALSE   2012 193      43  24   0   0
FALSE   2013   0       0  45  36   0
FALSE   2014 178       0  36   0   0
FALSE   2015   0       0  56 118   0
FALSE   2016 422     176  81   0   0
FALSE   2017   0       0 107 170  76
FALSE   2018 412       0 141   0   0
FALSE   2019   0       0 103  91  70
FALSE   2021   0       0 101 111  90
FALSE
FALSE $‘Yes, ’NA’ occured 100869 times in the column ’common_name’.‘
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA  NBS
FALSE   1982    0       0  925    0    0
FALSE   1983   67       0  878    0    0
FALSE   1984    0       0  942  172    0
FALSE   1985    0       0  959    0    0
FALSE   1986  158       0  846    0    0
FALSE   1987    0       0  839  164    0
FALSE   1988    0       0 1032    0    0
FALSE   1989    0       0  840    0    0
```

```
FALSE   1990    0       0 1065  189    0
FALSE   1991  338       0  936    0    0
FALSE   1992    0       0  852    0    0
FALSE   1993    0       0  948  870    0
FALSE   1994 1076       0  880    0    0
FALSE   1995    0       0  909    0    0
FALSE   1996    0       0 1015  914    0
FALSE   1997 1295       0  919    0    0
FALSE   1998    0       0 1011    0    0
FALSE   1999    0       0  972 1094    0
FALSE   2000 1533       0 1167    0    0
FALSE   2001    0       0  945 1067    0
FALSE   2002 1899     863 1176    0    0
FALSE   2003    0       0 1572 1612    0
FALSE   2004 2251    1314 1478    0    0
FALSE   2005    0       0 1421 1614    0
FALSE   2006 2205       0 1600    0    0
FALSE   2007    0       0 1372 1817    0
FALSE   2008    0    1102 1399    0    0
FALSE   2009    0       0 1454 1732    0
FALSE   2010 2162    1389 1670    0  964
FALSE   2011    0       0 1544 1178    0
FALSE   2012 2445    1127 1686    0    0
FALSE   2013    0       0 1655 1290    0
FALSE   2014 2473       0 1537    0    0
FALSE   2015    0       0 1656 2327    0
FALSE   2016 2451    1267 1526    0    0
FALSE   2017    0       0 1593 1443  918
FALSE   2018 2136       0 1600    0    0
FALSE   2019    0       0 1508 1495  804
FALSE   2021    0       0 1351 1258  718
FALSE
FALSE $‘Yes, ’NA’ occured 35359 times in the column ’scientific_name’.‘
FALSE       srvy
FALSE year    AI BSSlope EBS GOA NBS
FALSE   1982   0       0  46   0   0
FALSE   1983 186       0  56   0   0
FALSE   1984   0       0 456 658   0
FALSE   1985   0       0 275   0   0
FALSE   1986 266       0 316   0   0
FALSE   1987   0       0 418 547   0
FALSE   1988   0       0 353   0   0
FALSE   1989   0       0 273   0   0
FALSE   1990   0       0 272 567   0
FALSE   1991 223       0 287   0   0
FALSE   1992   0       0 254   0   0
FALSE   1993   0       0 314 688   0
FALSE   1994 332       0 397   0   0
FALSE   1995   0       0 465   0   0
FALSE   1996   0       0 556 542   0
FALSE   1997 657       0 552   0   0
FALSE   1998   0       0 578   0   0
FALSE   1999   0       0 641 455   0
FALSE   2000 746       0 602   0   0
```

```
FALSE   2001   0       0 575 353   0
FALSE   2002 483     189 550   0   0
FALSE   2003   0       0 590 558   0
FALSE   2004 574     116 628   0   0
FALSE   2005   0       0 619 584   0
FALSE   2006 273       0 654   0   0
FALSE   2007   0       0 632 504   0
FALSE   2008   0      86 639   0   0
FALSE   2009   0       0 668 418   0
FALSE   2010 520      24 656   0 219
FALSE   2011   0       0 590 445   0
FALSE   2012 472      65 651   0   0
FALSE   2013   0       0 662 336   0
FALSE   2014 435       0 634   0   0
FALSE   2015   0       0 657 556   0
FALSE   2016 294     110 685   0   0
FALSE   2017   0       0 651 406 266
FALSE   2018 338       0 690   0   0
FALSE   2019   0       0 633 266 267
FALSE   2021   0       0 604 284 272
FALSE
FALSE $‘Yes, ’NA’ occured 113703 times in the column ’cpue_noha’.‘
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA  NBS
FALSE   1982    0       0  821    0    0
FALSE   1983  200       0 1075    0    0
FALSE   1984    0       0 1601 1029    0
FALSE   1985    0       0 1410    0    0
FALSE   1986  239       0 1005    0    0
FALSE   1987    0       0 1192  332    0
FALSE   1988    0       0 1224    0    0
FALSE   1989    0       0 1383    0    0
FALSE   1990    0       0 1550  936    0
FALSE   1991  469       0 1912    0    0
FALSE   1992    0       0 1808    0    0
FALSE   1993    0       0 2043 1277    0
FALSE   1994  848       0 2083    0    0
FALSE   1995    0       0 2089    0    0
FALSE   1996    0       0 2098  733    0
FALSE   1997 1431       0 2291    0    0
FALSE   1998    0       0 2292    0    0
FALSE   1999    0       0 2018  559    0
FALSE   2000 1949       0 1876    0    0
FALSE   2001    0       0 1236  595    0
FALSE   2002 2225     327 1417    0    0
FALSE   2003    0       0 1346 1002    0
FALSE   2004 2415     361 1405    0    0
FALSE   2005    0       0 1186 1161    0
FALSE   2006 2682       0 1258    0    0
FALSE   2007    0       0 1435 1569    0
FALSE   2008    0     386 1490    0    0
FALSE   2009    0       0 1499 1320    0
FALSE   2010 3635     324 1653    0  875
FALSE   2011    0       0 1510 1375    0
```

```
FALSE   2012 3844     366 1521    0    0
FALSE   2013    0       0 1530 1051    0
FALSE   2014 3385       0 1442    0    0
FALSE   2015    0       0 1478 1836    0
FALSE   2016 3346     454 1535    0    0
FALSE   2017    0       0 1423 1236  719
FALSE   2018 2763       0 1750    0    0
FALSE   2019    0       0 1419  829  675
FALSE   2021    0       0 1315 1678  648
FALSE
FALSE $‘Yes, ’NA’ occured 113703 times in the column ’cpue_nokm2’.‘
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA  NBS
FALSE   1982    0       0  821    0    0
FALSE   1983  200       0 1075    0    0
FALSE   1984    0       0 1601 1029    0
FALSE   1985    0       0 1410    0    0
FALSE   1986  239       0 1005    0    0
FALSE   1987    0       0 1192  332    0
FALSE   1988    0       0 1224    0    0
FALSE   1989    0       0 1383    0    0
FALSE   1990    0       0 1550  936    0
FALSE   1991  469       0 1912    0    0
FALSE   1992    0       0 1808    0    0
FALSE   1993    0       0 2043 1277    0
FALSE   1994  848       0 2083    0    0
FALSE   1995    0       0 2089    0    0
FALSE   1996    0       0 2098  733    0
FALSE   1997 1431       0 2291    0    0
FALSE   1998    0       0 2292    0    0
FALSE   1999    0       0 2018  559    0
FALSE   2000 1949       0 1876    0    0
FALSE   2001    0       0 1236  595    0
FALSE   2002 2225     327 1417    0    0
FALSE   2003    0       0 1346 1002    0
FALSE   2004 2415     361 1405    0    0
FALSE   2005    0       0 1186 1161    0
FALSE   2006 2682       0 1258    0    0
FALSE   2007    0       0 1435 1569    0
FALSE   2008    0     386 1490    0    0
FALSE   2009    0       0 1499 1320    0
FALSE   2010 3635     324 1653    0  875
FALSE   2011    0       0 1510 1375    0
FALSE   2012 3844     366 1521    0    0
FALSE   2013    0       0 1530 1051    0
FALSE   2014 3385       0 1442    0    0
FALSE   2015    0       0 1478 1836    0
FALSE   2016 3346     454 1535    0    0
FALSE   2017    0       0 1423 1236  719
FALSE   2018 2763       0 1750    0    0
FALSE   2019    0       0 1419  829  675
FALSE   2021    0       0 1315 1678  648
FALSE
FALSE $‘Yes, ’NA’ occured 48632 times in the column ’bottom_temperature_c’.‘
```

```
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA
FALSE   1982    0       0  121    0
FALSE   1983  312       0   49    0
FALSE   1984    0       0  303 4348
FALSE   1985    0       0 1135    0
FALSE   1986 3198       0  268    0
FALSE   1987    0       0   87 6324
FALSE   1988    0       0   65    0
FALSE   1989    0       0  292    0
FALSE   1990    0       0 1714 5811
FALSE   1991 4241       0  893    0
FALSE   1992    0       0  918    0
FALSE   1993    0       0 1059 1024
FALSE   1994  127       0 2595    0
FALSE   1995    0       0  608    0
FALSE   1996    0       0  358 1917
FALSE   1997  210       0  631    0
FALSE   1998    0       0  771    0
FALSE   1999    0       0  856  426
FALSE   2000  312       0  770    0
FALSE   2001    0       0  501  756
FALSE   2002    0     181  307    0
FALSE   2003    0       0  256  309
FALSE   2004  155       0  301    0
FALSE   2005    0       0  908  482
FALSE   2006   31       0  160    0
FALSE   2007    0       0    0  869
FALSE   2008    0      65    0    0
FALSE   2009    0       0    0  277
FALSE   2010  141     127    0    0
FALSE   2011    0       0    0  102
FALSE   2012  373       0   57    0
FALSE   2014   51       0    0    0
FALSE   2015    0       0    0  370
FALSE   2016   45       0    0    0
FALSE   2017    0       0    0   27
FALSE   2021    0       0    0   38
FALSE
FALSE $‘Yes, ’NA’ occured 19709 times in the column ’surface_temperature_c’.‘
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA
FALSE   1983    0       0   76    0
FALSE   1984    0       0    0  960
FALSE   1985    0       0  703    0
FALSE   1986  356       0  118    0
FALSE   1987    0       0 2995  399
FALSE   1988    0       0  166    0
FALSE   1989    0       0  311    0
FALSE   1990    0       0   98 2694
FALSE   1991 1920       0  744    0
FALSE   1992    0       0  420    0
FALSE   1993    0       0  120  763
FALSE   1994  108       0   36    0
```

```
FALSE   1996    0       0    0 1710
FALSE   1997  210       0   25    0
FALSE   1998    0       0  104    0
FALSE   1999    0       0   71   13
FALSE   2000   70       0  202    0
FALSE   2001    0       0  253   84
FALSE   2002    0      46  115    0
FALSE   2003    0       0    0  253
FALSE   2004   31       0  252    0
FALSE   2005    0       0  104  345
FALSE   2006    0       0  160    0
FALSE   2007    0       0   34   44
FALSE   2008    0      65    0    0
FALSE   2009    0       0    0  247
FALSE   2010  141      71    0    0
FALSE   2011    0       0    0  102
FALSE   2012  415       0   57    0
FALSE   2013    0       0    0   12
FALSE   2014   51       0    0    0
FALSE   2015    0       0    0  394
FALSE   2016  193     120    0    0
FALSE   2017    0       0    0   57
FALSE   2018  157       0    0    0
FALSE   2019    0       0   31  270
FALSE   2021    0       0   27  186
FALSE
FALSE [1] "------------- Are there any '' in these columns? -------------"
FALSE [1] "------------- Are there any '0' in these columns? -------------"
FALSE $`Yes, '0' occured 113703 times in the column 'count'.`
FALSE       srvy
FALSE year     AI BSSlope  EBS  GOA  NBS
FALSE   1982    0       0  821    0    0
FALSE   1983  200       0 1075    0    0
FALSE   1984    0       0 1601 1029    0
FALSE   1985    0       0 1410    0    0
FALSE   1986  239       0 1005    0    0
FALSE   1987    0       0 1192  332    0
FALSE   1988    0       0 1224    0    0
FALSE   1989    0       0 1383    0    0
FALSE   1990    0       0 1550  936    0
FALSE   1991  469       0 1912    0    0
FALSE   1992    0       0 1808    0    0
FALSE   1993    0       0 2043 1277    0
FALSE   1994  848       0 2083    0    0
FALSE   1995    0       0 2089    0    0
FALSE   1996    0       0 2098  733    0
FALSE   1997 1431       0 2291    0    0
FALSE   1998    0       0 2292    0    0
FALSE   1999    0       0 2018  559    0
FALSE   2000 1949       0 1876    0    0
FALSE   2001    0       0 1236  595    0
FALSE   2002 2225     327 1417    0    0
FALSE   2003    0       0 1346 1002    0
FALSE   2004 2415     361 1405    0    0
```

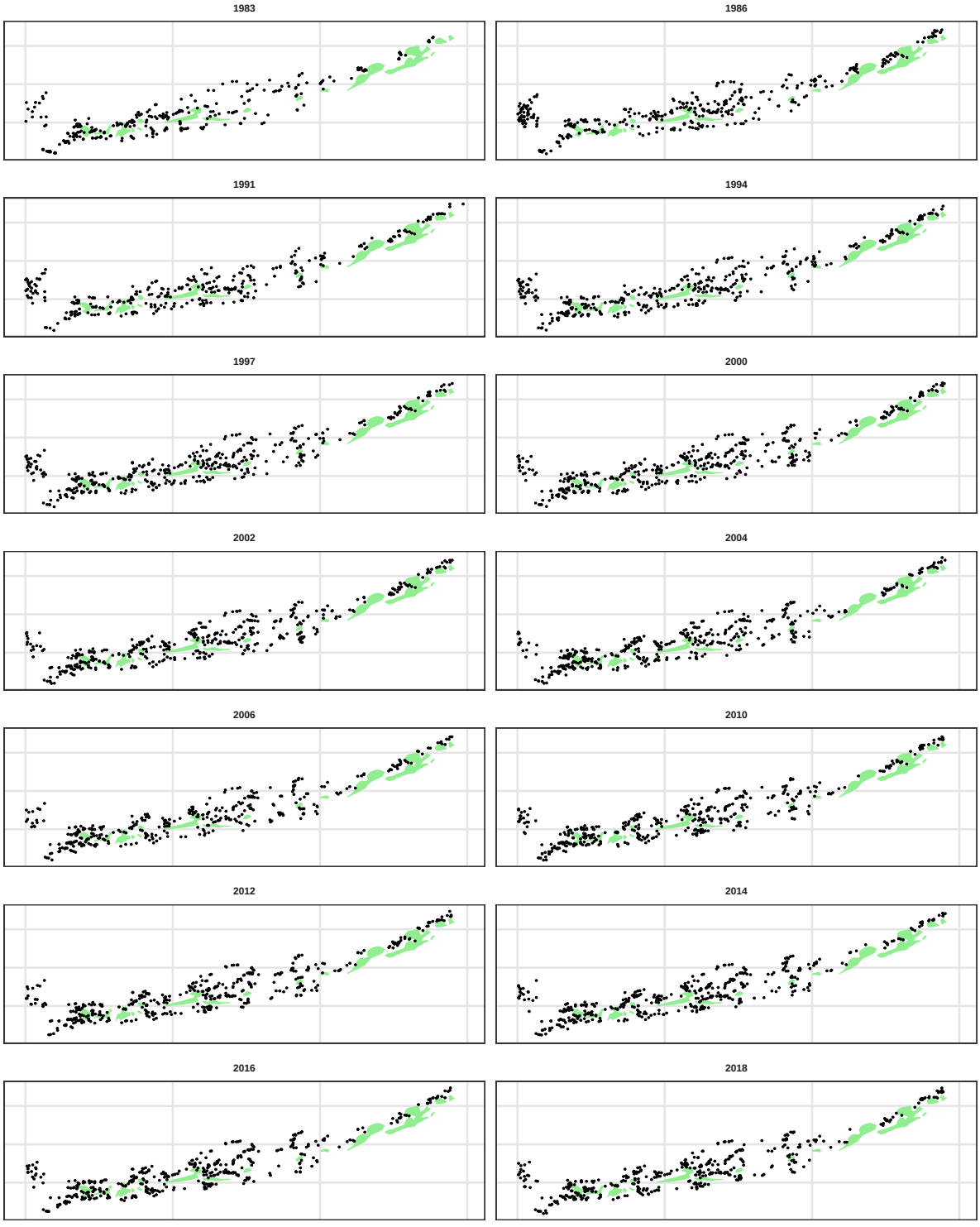
```
FALSE   2005    0       0 1186 1161    0
FALSE   2006 2682       0 1258    0    0
FALSE   2007    0       0 1435 1569    0
FALSE   2008    0     386 1490    0    0
FALSE   2009    0       0 1499 1320    0
FALSE   2010 3635     324 1653    0  875
FALSE   2011    0       0 1510 1375    0
FALSE   2012 3844     366 1521    0    0
FALSE   2013    0       0 1530 1051    0
FALSE   2014 3385       0 1442    0    0
FALSE   2015    0       0 1478 1836    0
FALSE   2016 3346     454 1535    0    0
FALSE   2017    0       0 1423 1236  719
FALSE   2018 2763       0 1750    0    0
FALSE   2019    0       0 1419  829  675
FALSE   2021    0       0 1315 1678  648
FALSE
FALSE $‘Yes, ’0’ occured 3659 times in the column ’bottom_temperature_c’.‘
FALSE       srvy
FALSE year   EBS NBS
FALSE   1982 110   0
FALSE   1983 129   0
FALSE   1984 295   0
FALSE   1985  38   0
FALSE   1986  28   0
FALSE   1987  27   0
FALSE   1988  45   0
FALSE   1990  22   0
FALSE   1991  40   0
FALSE   1992 156   0
FALSE   1993  35   0
FALSE   1994 231   0
FALSE   1995  46   0
FALSE   1996  76   0
FALSE   1999 294   0
FALSE   2000  66   0
FALSE   2001  96   0
FALSE   2002  34   0
FALSE   2004  21   0
FALSE   2005  74   0
FALSE   2006  80   0
FALSE   2007 195   0
FALSE   2008 157   0
FALSE   2009 118   0
FALSE   2010 138  29
FALSE   2011  50   0
FALSE   2012 466   0
FALSE   2013 168   0
FALSE   2014  95   0
FALSE   2015 141   0
FALSE   2017  65  37
FALSE   2019   0  33
FALSE   2021  24   0
FALSE
```

```
FALSE $‘Yes, ’0’ occured 116 times in the column ’surface_temperature_c’.‘
FALSE       srvy
FALSE year   EBS
FALSE   1982  13
FALSE   1999  73
FALSE   2008  30
```

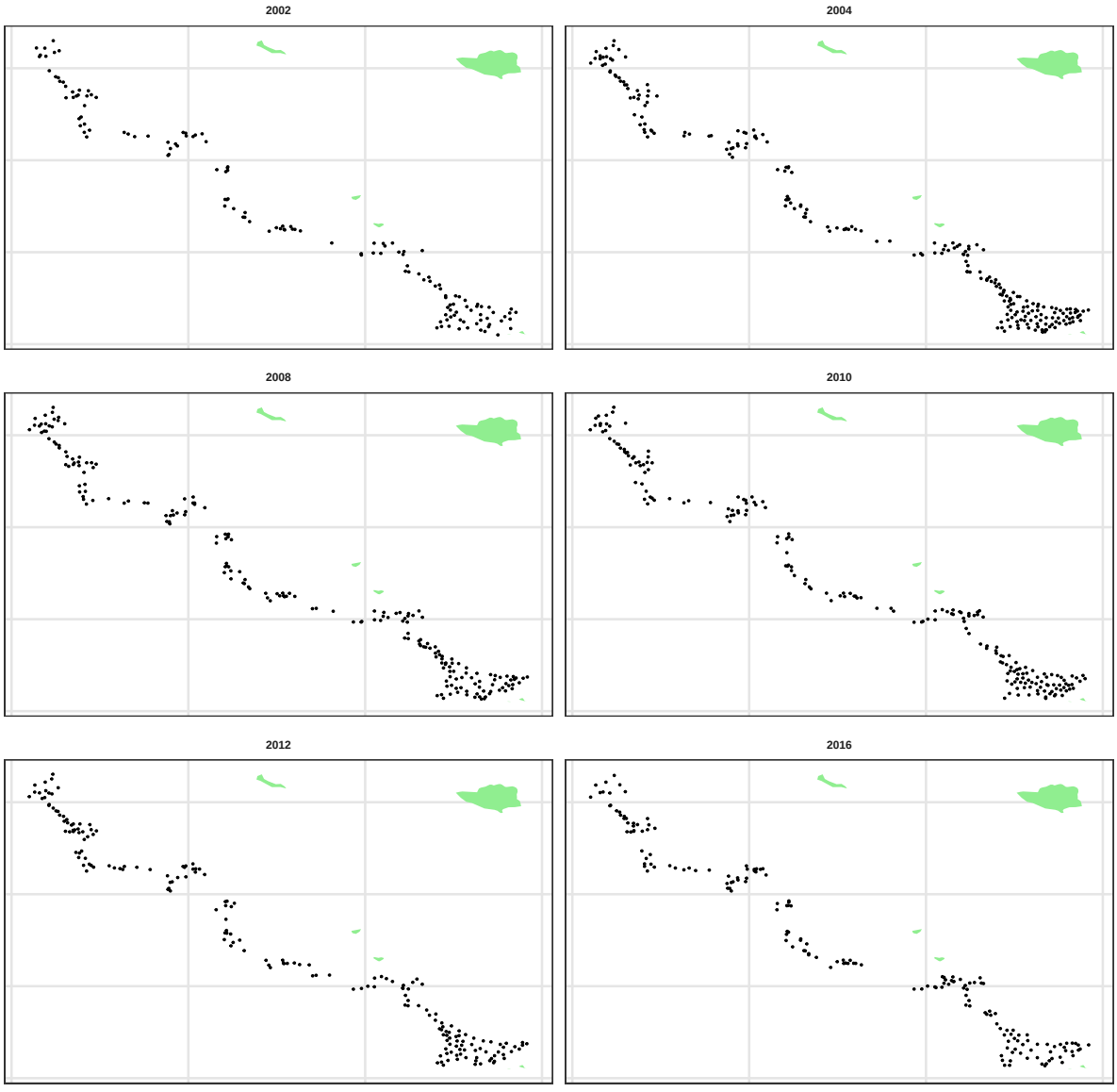
What stations for each year and survey look like

```
for (i in 1:length(unique(data_new$srvy))) {
  print(plot_stations(dat = data_new, srvy0 = unique(data_new$srvy)[i]))
}
```

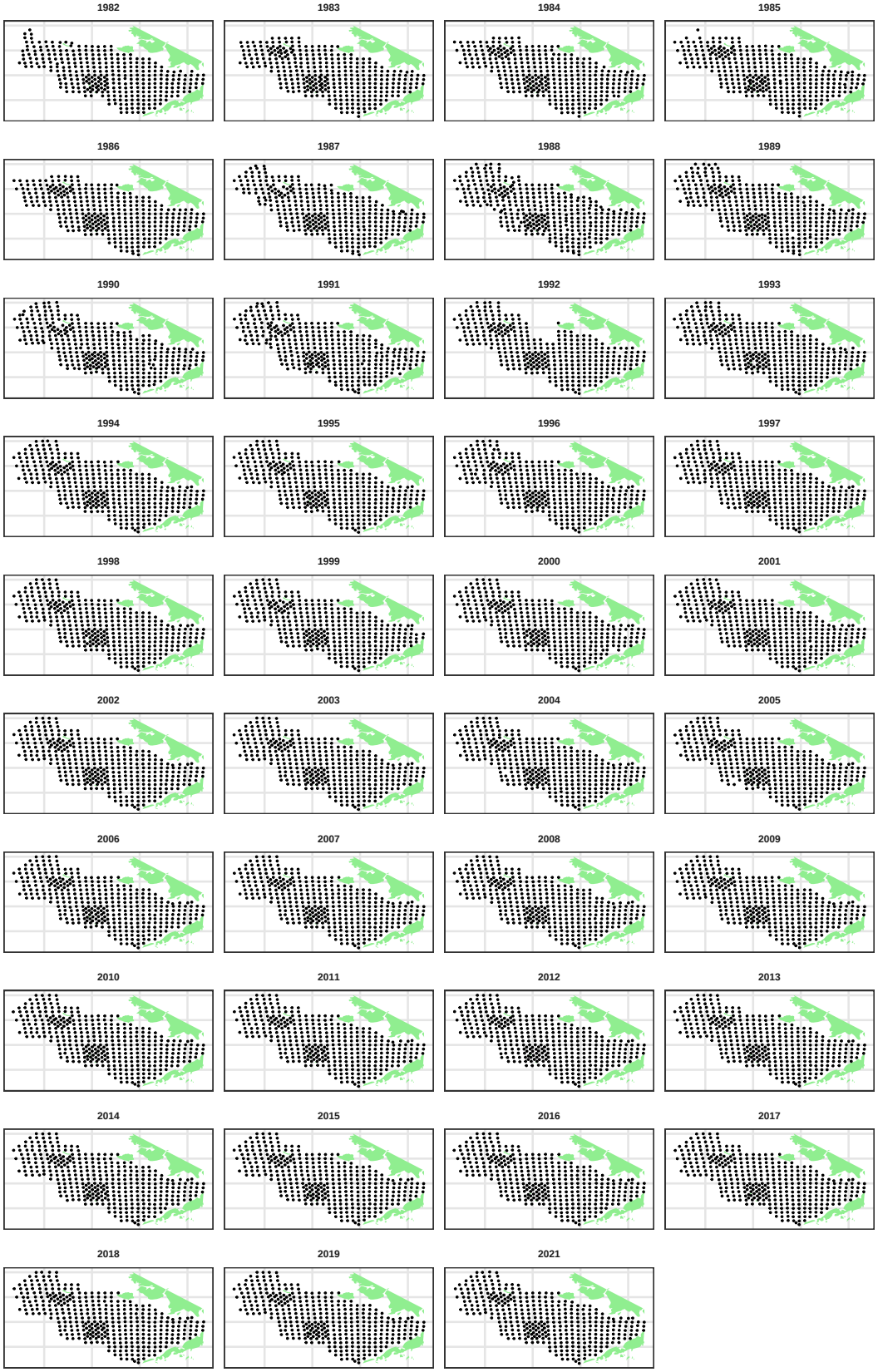
AI


1983
1986
1991
1994
1997
2000
2002
2004
2006
2010
2012
2014
2016
2018

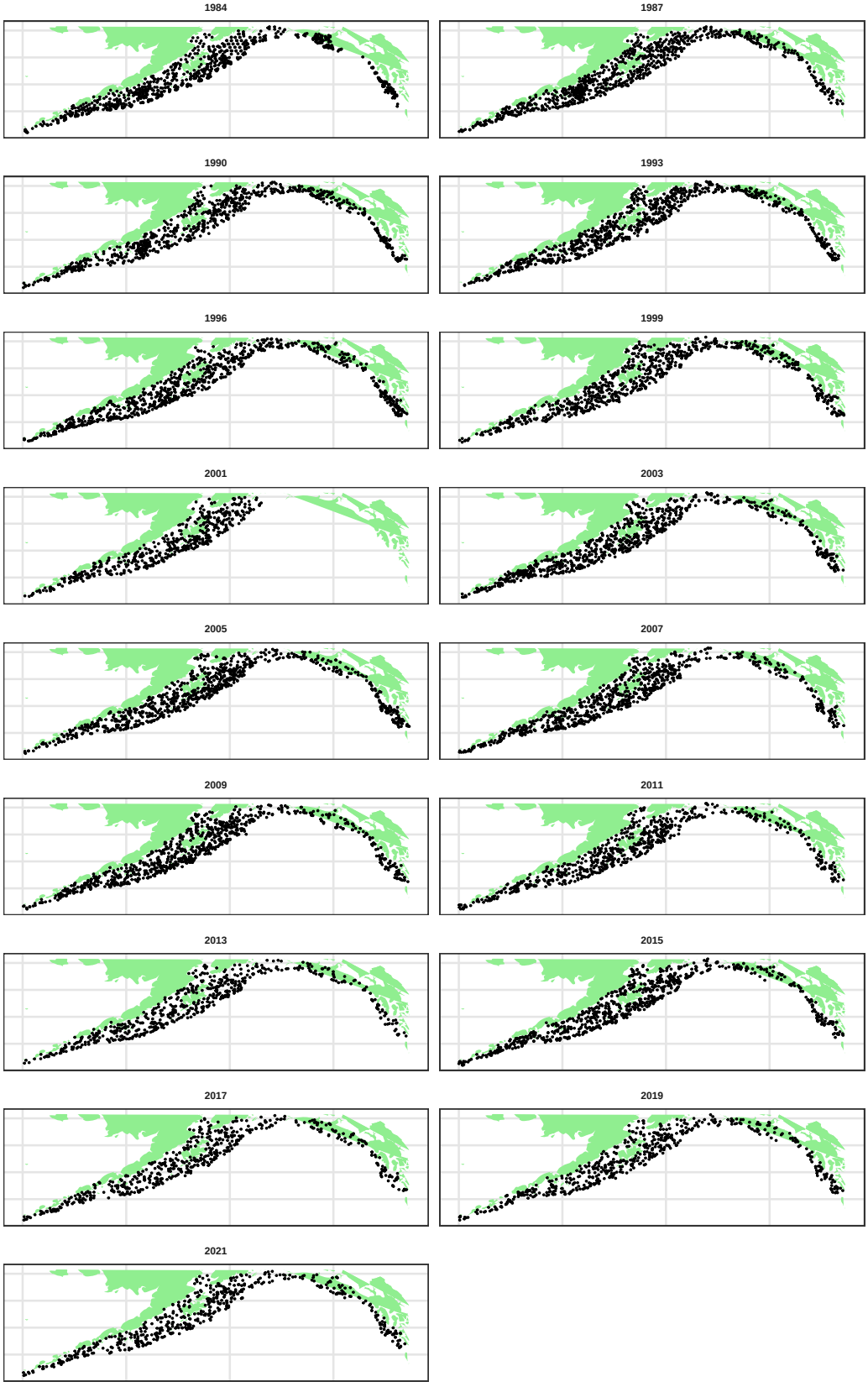
BSSlope



EBS


1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2021

GOA


1984
1987
1990
1993
1996
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
2021

NBS