

Anonymization

In this section we'll work through variables with potentially identifiable information and either exclude/anonymize them.

Identifying Variables for Anonymization

Continuous Variables

Survey Variables

Variables with numeric or integer inputs can be identifiable using publicly available 990 data. They are presented in the table below.

Variable	Description
Reserves__Est	Dollar Value of Financial Reserves
CARES_Rcv__Est	Dollar Value of CARES Funding
Finance_Rev_GovtMain	Dollar Value of Revenue from Government Agencies
Finance_Rev_Prtcptnt	Dollar Value of Revenue from Self-Paying Participants
Finance_Rev_Govt3rdParty	Dollar Value of Revenue from Government as Third-Party Payer
Finance_Rev_IndvDon	Dollar Value of Revenue from Individual Donations
Finance_Rev_Gift	Dollar Value of Revenue from In-Kind Gifts
Finance_Rev_Grnt	Dollar Value of Revenue from Private Foundations/Corporate Grants
Finance_Rev_Spnsr	Dollar Value of Revenue from Corporate Sponsorships
Finance_Rev_Oth	

We anonymize by randomly sampling and adding a small amount of noise to the data.

```
cols_to_anonymize <- c(  
  "CARES_Rcv__Est",  
  "Finance_Rev_GovtMain",  
  "Finance_Rev_Prtcptnt",  
  "Finance_Rev_Govt3rdParty",  
  "Finance_Rev_IndvDon",
```

```

    "Finance_Rev_Gift",
    "Finance_Rev_Grnt",
    "Finance_Rev_Spnsr",
    "Finance_Rev_Oth"
  )

  # Create anonymization function
  anon_continuous <- function(var){
    var <- var * (1 + runif(n=1, min=-0.05, max=0.05))
    var <- var * (1 + rnorm(n=1, sd=0.05))
    return(var)
  }

  # Anonymize Continuous Variables
  survey_anon_df <- survey_census_df %>%
    dplyr::mutate(
      dplyr::across(
        tidyselect::all_of(cols_to_anonymize),
        ~ anon_continuous(.x)
      )
    )

```

Census Variables

Similarly, census metadata at tract and CBSA levels can be used to identify the specific location of a nonprofit, leading to identification. We apply the same anonymization process to these columns.

```

# Retrieve Census variables for anonymization
census_cols_to_anonymize <-
  c(
    "total_population_County",
    "housing_units_County",
    "occupied_County",
    "vacant_County",
    "renter_occ_County",
    "white_perc_County",
    "black_perc_County",
    "asian_perc_County",
    "hawaiian_perc_County",

```

"american_alaskan_perc_County",
"two_or_more_perc_County",
"other_perc_County",
"rural_perc_County",
"bachelors_perc_County",
"hispanic_perc_County",
"poverty_perc_County",
"unemployment_County",
"turnover_perc_County",
"med_family_income_adj_County",
"med_gross_rent_adj_County",
"med_household_income_adj_County",
"median_value_adj_County",
"total_population_Tract",
"housing_units_Tract",
"occupied_Tract",
"vacant_Tract",
"renter_occ_Tract",
"white_perc_Tract",
"black_perc_Tract",
"asian_perc_Tract",
"hawaiian_perc_Tract",
"american_alaskan_perc_Tract",
"two_or_more_perc_Tract",
"other_perc_Tract",
"rural_perc_Tract",
"bachelors_perc_Tract",
"hispanic_perc_Tract",
"poverty_perc_Tract",
"unemployment_Tract",
"turnover_perc_Tract",
"med_family_income_adj_Tract",
"med_gross_rent_adj_Tract",
"med_household_income_adj_Tract",
"median_value_adj_Tract",
"total_population_CBSA",
"housing_units_CBSA",
"occupied_CBSA",
"vacant_CBSA",
"renter_occ_CBSA",
"unemployment_CBSA",

```

    "med_family_income_adj_CBSA",
    "med_gross_rent_adj_CBSA",
    "med_household_income_adj_CBSA",
    "median_value_adj_CBSA"
  )

# Run anonymization
# Anonymize Continuous Variables
survey_anon_df <- survey_anon_df %>%
  dplyr::mutate(
    dplyr::across(
      tidyselect::all_of(census_cols_to_anonymize),
      ~ anon_continuous(.x)
    )
  )

```

Selecting Columns for Release in Final Data

With anonymization complete, we can remove columns with identifiable information and information not relevant to the survey before release.

```

release_cols <-
  c(
    "PrgSrvc_IncrNum",
    "PrgSrvc_DcrsNum",
    "PrgSrvc_Suspend",
    "PrgSrvc_IncrSrvc",
    "PrgSrvc_DcrsSrvc",
    "PrgSrvc_NewOffc",
    "PrgSrvc_ClsdOffc",
    "PrgSrvc_IncrArea",
    "PrgSrvc_DcrsArea",
    "PrgSrvc_IncrFee",
    "PrgSrvc_DcrsFee",
    "PrgSrvc_ShiftOnline",
    "PrgSrvc_AddOnline",
    "PrgSrvc_Oth",
    "PrgSrvc_OthChng_Text",
    "PplSrv_NumServed",
  )

```

"PplSrv_NumWait",
"PplSrv_NumServed_NA_X",
"PplSrv_NumWait_NA_X",
"Dmnd_NxtYear",
"Staff_Fulltime_2021",
"Staff_Parttime_2021",
"Staff_Boardmmbr_2021",
"Staff_RegVlntr_2021",
"Staff_EpsdVlntr_2021",
"Staff_AmerVlntr_2021",
"Staff_PdCnslt_2021",
"Staff_Other_Text_2021",
"Staff_Other_Est_2021",
"Staff_Fulltime_2022",
"Staff_Parttime_2022",
"Staff_Boardmmbr_2022",
"Staff_RegVlntr_2022",
"Staff_EpsdVlntr_2022",
"Staff_AmerVlntr_2022",
"Staff_PdCnslt_2022",
"Staff_Other_Text_2022",
"Staff_Other_Est_2022",
"Staff_Fulltime_NA",
"Staff_Parttime_NA",
"Staff_Boardmmbr_NA",
"Staff_RegVlntr_NA",
"Staff_EpsdVlntr_NA",
"Staff_AmerVlntr_NA",
"Staff_PdCnslt_NA",
"Staff_Other_Text_NA",
"Staff_Other_Est_NA",
"VolImportance",
"DonImportance",
"FndRaise_DvlpVirtual",
"FndRaise_IncrExp",
"FndRaise_DcrsExp",
"FndRaise_IncStaff",
"FndRaise_DcrsStaff",
"FndRaise_IncrCnslt",
"FndRaise_DcrsCnslt",
"FndRaise_IncrVolntr",

"FndRaise_DcrsVolntr",
"FndRaise_Othr",
"FndRaise_Text",
"FndRaise_MajGift_Amt",
"FndRaise_Overall_Chng",
"FndRaise_Cashbelow250_Chng",
"FndRaise_Cashabove250_Chng",
"FndRaise_MajGift_Chng",
"FndRaise_Corp-Found_Grnt_Chng",
"FndRaise_RstrGift_Chng",
"FndRaise_UnrstrGift_Chng",
"FndRaise_DnrBlw250",
"FndRaise_DnrAbv250",
"FndRaise_LocGvtGrnt_Seek",
"FndRaise_StateGvtGrnt_Seek",
"FndRaise_FedGvtGrnt_Seek",
"FndRaise_LocGvtCntrct_Seek",
"FndRaise_StateGvtCntrct_Seek",
"FndRaise_FedGvtCntrct_Seek",
"FndRaise_PFGnt_Seek",
"FndRaise_CFGnt_Seek",
"FndRaise_DAF_Seek",
"FndRaise_Corp-Found_Grnt_Seek",
"FndRaise_UntdWy_Seek",
"FndRaise_CombFedCmpgn_Seek",
"FndRaise_OthrGvngPrgrm_Seek",
"FndRaise_LocGvtGrnt_Rcv",
"FndRaise_StateGvtGrnt_Rcv",
"FndRaise_FedGvtGrnt_Rcv",
"FndRaise_LocGvtCntrct_Rcv",
"FndRaise_StateGvtCntrct_Rcv",
"FndRaise_FedGvtCntrct_Rcv",
"FndRaise_PFGnt_Rcv",
"FndRaise_CFGnt_Rcv",
"FndRaise_DAF_Rcv",
"FndRaise_Corp-Found_Grnt_Rcv",
"FndRaise_UntdWy_Rcv",
"FndRaise_CombFedCmpgn_Rcv",
"FndRaise_OthrGvngPrgrm_Rcv",
"Finance_Rev_GovtMain",
"Finance_Rev_Prtcpt",

"Finance_Rev_Govt3rdParty",
"Finance_Rev_IndvDon",
"Finance_Rev_Gift",
"Finance_Rev_Grnt",
"Finance_Rev_Spnsr",
"Finance_Rev_Oth",
"Finance_Rev_Oth_Text",
"Time_FirstClick_Finances",
"Time_LastClick_Finances",
"Time_PageSubmit_Finances",
"Time_ClickCount_Finances",
"Reserves_Est",
"Reserves_NA_X",
"CARES_Rcv",
"CARES_Rcv_Est",
"FinanceChng_Reserves",
"FinanceChng_Borrow",
"FinanceChng_DcrsBnft",
"FinanceChng_IncrBnft",
"FinanceChng_IncrExp",
"FinanceChng_DcrsExp",
"FinanceChng_Oth",
"FinanceChng_Text",
"LeadershipChng_RetCEO",
"LeadershipChng_RsgnCEO",
"LeadershipChng_TrmnCEO",
"LeadershipChng_HireCEO",
"LeadershipChng_IntrmCEO",
"LeadershipChng_ChngBC",
"LeadershipChng_LostBoardMem",
"LeadershipChng_RplcBoardMem",
"LeadershipChng_AddBoardMem",
"LeadershipChng_Oth",
"LeadershipChng_Oth_Text",
"CEOrace_AAPI",
"CEOrace_Black",
"CEOrace_Hisp",
"CEOrace_NativeAm",
"CEOrace_White",
"CEOrace_Oth",
"CEOrace_Text",

"CEOgender_Man",
"CEOgender_Woman",
"CEOgender_Trans",
"CEOgender_NB",
"CEOgender_Oth",
"CEOgender_specify",
"BChairrace_AAPIO",
"BChairrace_Black",
"BChairrace_Hisp",
"BChairrace_NativeAm",
"BChairrace_White",
"BChairrace_Oth",
"BChairrace_Text",
"BChairgender_Man",
"BChairgender_Woman",
"BChairgender_Trans",
"BChairgender_NB",
"BChairgender_Oth",
"BChairgender_specify",
"ExtAffairs_GenEd",
"ExtAffairs_Media",
"ExtAffairs_Advocacy",
"ExtAffairs_GovtRs",
"ExtAffairs_DiscGovtGrnt",
"ExtAffairs_InfoReq",
"ExtAffairs_Testify",
"ExtAffairs_Lobby",
"ExtAffairs_OrgPrtst",
"ExtAffairs_Mobilize",
"ExtAffairs_Petition",
"ExtAffairs_VoteReg",
"ExtAffairs_VoteEd",
"PrimaryConcern",
"CEOrace_Bi",
"BChairrace_Bi",
"State",
"StateAbbr",
"County",
"City",
"total_population_County",
"housing_units_County",

"occupied_County",
"vacant_County",
"renter_occ_County",
"white_perc_County",
"black_perc_County",
"asian_perc_County",
"hawaiian_perc_County",
"american_alaskan_perc_County",
"two_or_more_perc_County",
"other_perc_County",
"rural_perc_County",
"bachelors_perc_County",
"hispanic_perc_County",
"poverty_perc_County",
"unemployment_County",
"turnover_perc_County",
"med_family_income_adj_County",
"med_gross_rent_adj_County",
"med_household_income_adj_County",
"median_value_adj_County",
"total_population_Tract",
"housing_units_Tract",
"occupied_Tract",
"vacant_Tract",
"renter_occ_Tract",
"white_perc_Tract",
"black_perc_Tract",
"asian_perc_Tract",
"hawaiian_perc_Tract",
"american_alaskan_perc_Tract",
"two_or_more_perc_Tract",
"other_perc_Tract",
"rural_perc_Tract",
"bachelors_perc_Tract",
"hispanic_perc_Tract",
"poverty_perc_Tract",
"unemployment_Tract",
"turnover_perc_Tract",
"med_family_income_adj_Tract",
"med_gross_rent_adj_Tract",
"med_household_income_adj_Tract",

```

    "median_value_adj_Tract",
    "cbsa_title_CBSA",
    "metro_CBSA",
    "total_population_CBSA",
    "housing_units_CBSA",
    "occupied_CBSA",
    "vacant_CBSA",
    "renter_occ_CBSA",
    "unemployment_CBSA",
    "med_family_income_adj_CBSA",
    "med_gross_rent_adj_CBSA",
    "med_household_income_adj_CBSA",
    "median_value_adj_CBSA"
  )

survey_release_df <- survey_anon_df %>%
  dplyr::select(
    tidyselect::all_of(release_cols)
  )

```

Saving Release Data Set

```
# Save anonymized data
setwd("Y:/CNP/Generosity Commission/")
readr::write_csv(
  survey_anon_df,
  "DATA-PREP/02-data-intermediate/02-wave-two/wave-02-data-intermediate-anonymized.csv"
)

# Save release data
setwd("Y:/CNP/Generosity Commission/")
readr::write_csv(
  survey_release_df,
  "DATA-PREP/03-data-final/Public-use/nptrends-puf-wave-02.csv"
)
```