

1) DATA STEPS

```
/* EXAMPLE: Some Swedish prime ministers */

/* The INPUT statement */

data ministers1;
input year minister $; /* $ denotes character variables */
datalines; /* "cards;" can be written instead of "datalines;" */
2022 Ulf Kristersson
2021 Magdalena Andersson
2014 Stefan Löfven
2006 Fredrik Reinfeldt
1996 Göran Persson
1994 Ingvar Carlsson
1991 Carl Bildt
1986 Ingvar Carlsson
1982 Olof Palme
;
run;

/* Mark/select the code, from "data" to "run;", and press the running
figure or F3 to run the selected code */

/* Only first name is included in the created dataset */

data ministers2;
input year minister & $; /* & tells SAS to use two whitespace
                           characters to signal end of character */
datalines;
2022 Ulf Kristersson
2021 Magdalena Andersson
2014 Stefan Löfven
2006 Fredrik Reinfeldt
1996 Göran Persson
1994 Ingvar Carlsson
1991 Carl Bildt
1986 Ingvar Carlsson
1982 Olof Palme
;
run;
/* Includes white space but not all characters */
```

```
/* There are two ways to set the length of a variable: */

data ministers3;
length minister $ 20; /* Sets the length of the variable */
input year minister & $;
datalines;
2022 Ulf Kristersson
2021 Magdalena Andersson
2014 Stefan Löfven
2006 Fredrik Reinfeldt
1996 Göran Persson
1994 Ingvar Carlsson
1991 Carl Bildt
1986 Ingvar Carlsson
1982 Olof Palme
;
run;

/* or */

data ministers3;
input year minister & $ 20.; /* Sets the length of the variable */
datalines;
2022 Ulf Kristersson
2021 Magdalena Andersson
2014 Stefan Löfven
2006 Fredrik Reinfeldt
1996 Göran Persson
1994 Ingvar Carlsson
1991 Carl Bildt
1986 Ingvar Carlsson
1982 Olof Palme
;
run;

/* Done! */
```

```

/* The INFILE statement */

/* Open the file Prime ministers.txt in a text editor */
/* Looks like tab separated data */

/* To find the location of the file in SAS Studio: Find the file in the
folder you uploaded it to, under Server Files and Folders - Files (Home).
Right click and choose Properties. Copy the path from Location
(Ctrl/Command+C) and paste it in your code (Ctrl/Command+V), with quotation
marks around it (single or double) */

data ministers4;
  infile '/home/username/SAS introduction/Data/Prime ministers.txt'
    dlm='09'x; /* Tab delimiter */
  input year minister & $20.;

run;
/* Heading row imported as data */
/* Swedish letters not imported properly */

data ministers4;
  infile '/home/username/SAS introduction/Data/Prime ministers.txt'
    dlm='09'x /* Tab delimiter */
    firstobs=2 /* First row to be imported */
    encoding='ISO-8859-1'; /* Support special characters */
  input year minister & $20.;

run;

/* Import a file with additional information after the last observation */
data ministers5;
  infile '/home/username/SAS introduction/Data/Prime ministers w
headings.txt'
    dlm='09'x /* Tab delimiter */
    firstobs=2 /* First row to be imported */
    obs = 10 /* Last row number to be imported */
    encoding='ISO-8859-1'; /* Support special characters */
  input year minister & $20.;

run;

```

```

/* PROC IMPORT */

proc import
  dbms=xlsx /* Denotes the type of file to be imported */
  datafile='/home/username/SAS introduction/Data/Prime
ministers.xlsx'
  out=minsters6
  replace;
run;

* 'replace' tells sas to overwrite if a file with this name already exists;

/* If SAS cannot read the variable names, default names like VAR2, VAR2,
etc. will be created automatically. */

/* SET statement */
/* Keep the "best" ministers file */
data ministers; /* Create a new data set */
  set ministers4; /* based on ministers4 */
  yearsago=2024-year; /* Create a new variable */
run;


/* IF, THEN and ELSE statements */

/* Calculate how many years each minister held the post */
data ministers7;
  set ministers;
  endyear=lag(year);
  period=endyear-year;
run;

data ministers7;
  set ministers7;
  if period = . then periodSoFar=2014-year;
  else periodSoFar=period;
run;

```

```

/* KEEP and DROP statements */

data yearonly;
    set ministers;
    keep year;           /* keep only selected variables */
run;

/* or */
data yearonly;
    set ministers (keep=year);
run;


data ministersonly;
    set ministers;
    drop year;           /* drop selected variables */
run;

/* or */
data ministersonly;
    set ministers (drop=year);
run;


/* WHERE statement */

data ministers2000;
    set ministers;
    where year>=2000;   /* keep only selected observations */
run;


/* OUTPUT and DELETE statements */

/* Create a data set only containing the ministers who sat for more than 4
years */
data ministers4years;
    set ministers7;
    if period>4 then output;
run;

/* or */
data ministers4years;
    set ministers7;
    if period<=4 then delete;
run;

```

```

/* MERGE statement */

/* Create new data set with political party */
data ministers8;
input year party $;
datalines;
2022 M
2021 S
2014 S
2006 M
1996 S
1994 S
1991 M
1986 S
1982 S
;
run;

/* Merge the original data set with the new one */
data ministers9;
    merge ministers ministers8;
    by year;
run;

/* "ERROR: BY variables are not properly sorted on data set
WORK.MINISTERS" */

/* Sort the data: */

proc sort data=ministers;
    by year;
run;

proc sort data=ministers8;
    by year;
run;

/* Then merge again: */

data ministers9;
    merge ministers ministers8;
    by year;
run;

```

```
/* CREATE PERMANENT DATA SETS */

/* First assign which directory/library to use with the libname statement
(the run; command is not needed): */

libname intro '/home/username/SAS introduction/Data';
/* Select a name for the folder/library to be used in SAS, "intro" above */

/* Then use the name of that folder as a suffix to your dataset: */
data intro.ministers;
set ministers;
run;

/* DELETE SAS WORK DATA SETS */

proc datasets lib=work nolist memtype=data kill;
run;
quit;
```