1 Computing for Chapter 2 of Box, Jenkins, & Reinsel.

To solve the homework exercises for this chapter, you should be able to

(1) Produce a *time series plot*, i.e. plot an observed series z_t, $t = 1, \ldots, n$, as a function of time t.

(2) Produce a lagged scatterplot, i.e. plot z_{t+h} vs. z_t for some lag h, typically $h = 1$ or 2, but not much more.

(3) Calculate and plot the sample autocovariance and autocorrelation, and plot the requisite bands to indicate significant sample autocorrelations.

(4) Compute and plot a sample spectrum (periodogram).

We will defer the spectrum to another day and only treat the first 3.

We carry this exercise out with the cow temperature data set. This is in the file "cow.dat" which you may download from the same place I got it:

> http://www-personal.buseco.monash.edu.au/ hyndman/forecasting/welcome.htm
>
> go to the "Data" link on the sidebar, and the cow data is under Exercise 2.4.

In section 2, we show how to do this with SAS. In section 3, we show how to do this with Splus.

2 Using SAS to do Time Series Plots and Plots of the Sample ACF (Autocorrelation Function).

I have already downloaded the data into a file "cow.dat".

I fired up the SAS system, and a number of windows appeared on my screen. To get the data into SAS, I typed the following lines into the SAS: PROGRAM EDITOR window:

```
filename mydata 'cow.dat';
data cow;
infile mydata;
input temp;
run;
```

(Actually, there are some 5 digit numbers already numbering the lines in the PROGRAM EDITOR window which I didn't show.) I learned about this method from the "Reading from an External File" at the New Zealand sasintro web site. NOTE HOW ALL SAS STATEMENTS END WITH A SEMICOLON. My data set will be called "cow" and it will have only one variable, "temp" (for temperature).

I then select "Submit" from the Locals menu (top of the PROGRAM EDITOR window in the bar) and the statements are executed. The SAS LOG window tells me there is an error because it can't find the data file cow.dat in the directory where I am running SAS, so I move the file into that directory. I then select "Recall Text" from the Locals menu and all 5 lines reappear. I do the "Submit" again and the SAS LOG window tells me 75 records were read. My cow data set (with the temp variable) is in SAS now.

I next want to do a time series plot (or time plot) of the data. I really don't know how to do this, so I select "Extendend Help" from the Help menu in the PRO-

GRAM EDITOR window and a new SAS System Help window pops up (you can fill your desktop in a hurry when you run SAS). I select "MODELING & ANALYSIS TOOLS" in the Help window after a little trial and error and seeing that there is an "Econometrics & Time Series" link from there, which I take. Then I select the "Time Series Viewer" link. Based on what I read there, I type "tsview data=cow var=temp;" in the PROGRAM EDITOR window and get an error message in the SAS LOG window. So I go to the "Toolbox: PROGRAM EDITOR" window (which is real skinny top to bottom) and has a little place where you type in commands and typed in "tsview data=cow var=temp" (no ";" at the end) and left mouse click on the check on the left end of this window. A funny looking error message window pops up that makes very little sense to me but I get the idea that I hadn't properly specified the ID variable in the tsview command.

So, I went back and redid the DATA step:

```
me mydata 'cow.dat';
data cow;
infile mydata;
input temp;
time=_N_;
run;
```

The SAS LOG window tells me I have a data set with 2 variables: temp and time. I go back and redo the "tsview data=cow var=temp" in the little Toolbox window and lo and behold a really cute little time plot of my data pops up. I want to get the plot out into some file so I somehow figure out how to use the File $\rightarrow$ Print Graph menu to "print" the graph as a GIF file, which you see below. (I haven't been able to reproduce that.) It turns up in my directory as "graph.gsf" and I change it to tsplot.gsf.

I get a plot of the sample autocorrelation function either by going to the View menu or clicking the unobvious second button on the right side of the window. The autocorrelation appears (vertically) on the left side of the next picture. (I changed the file name to acf.gsf (acf for "autocorrelation function"), something I do to avoid having the file clobbered by the next print). I also try to File → Print Data so as to get a printout of the sample, but that seems to hang up SAS. I guess that isn't supposed to be done. Anyway, I go back and find in the Options menu a "Correlation Probabilities" option that allows me to get some lines on my plots indicating which correlations are significantly nonzero.

The numerical values of the sample acf appear in the SAS: OUTPUT window (along with a lot of other crap). Here is the 1st few lines:

The SAS System 1

15:53 Tuesday, February 2, 1999

Autocorrelation Plots

TEMP

LAG	N	ACOV	LACF	ACF	UACF	ACF_PRB	LPACF	PACF	UPACF
0	75	90.923733333	0.000	1	0.000	0.0001	0.000	1	0.000
1	74	37.324672	-.231	.41	0.231	0.0004	-.231	.41	0.231
2	73	38.456810667	-.267	.42	0.267	0.0015	-.231	.31	0.231
3	72	36.824682667	-.301	.41	0.301	0.0071	-.231	.21	0.231
4	71	33.882154667	-.328	.37	0.328	0.0233	-.231	.13	0.231

```
  5   70      14.82016  -.350  .16  0.350  0.3520   -.231  -.2   0.231
```

The sample ACF appears under the column "ACF".

Getting the lagged scatterplot is even more fun. The only way I can figure out how to do it is to get a new variable in the data set which is the lagged value of temp. Back to the data step:

```
filename mydata 'cow.dat';
data cow;
infile mydata;
input temp;
time=_N_;
templag=lag1(temp);
run;
```

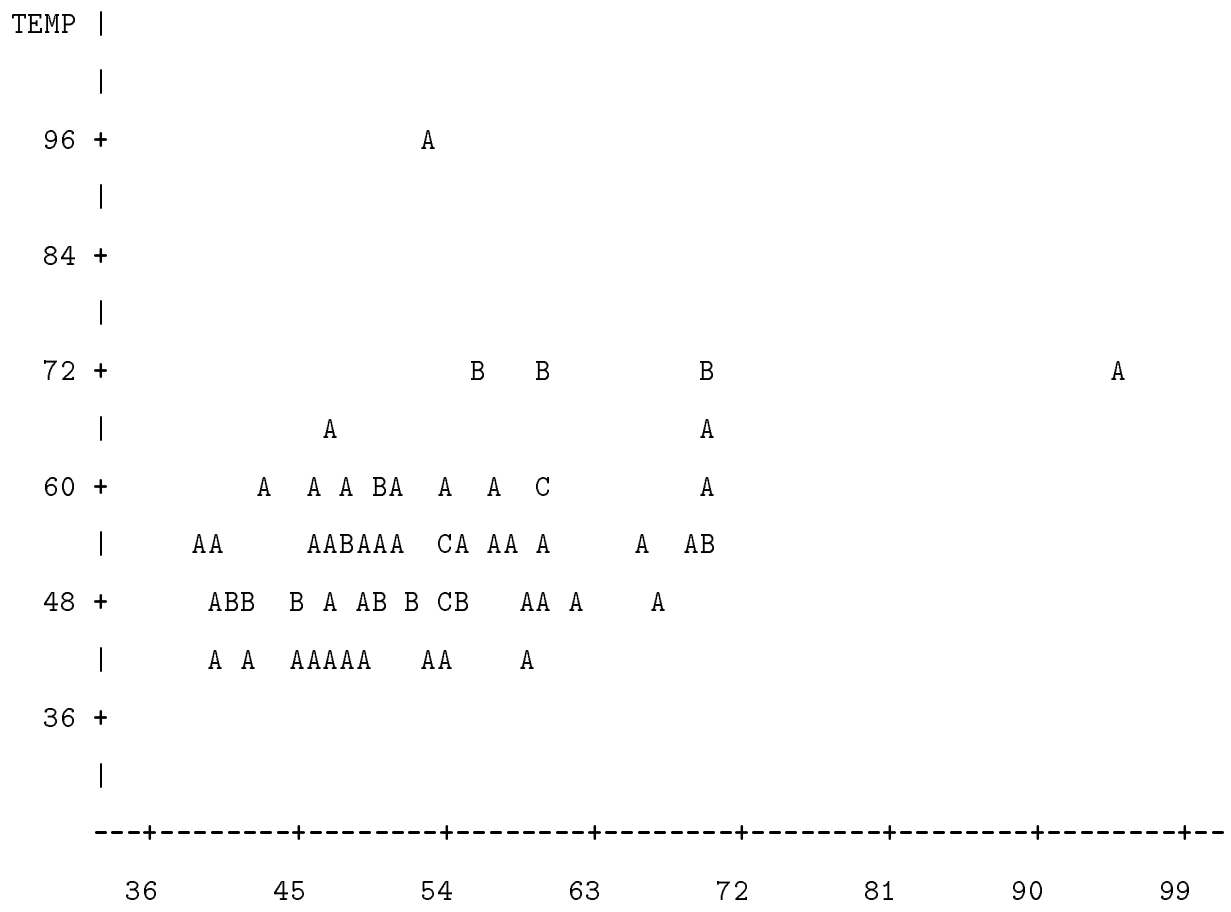
After submitting this, I invoked PROC PLOT to get the plot.

```
proc plot;
plot temp*templag;
run;
```

This produces a lousy line printer plot that appears in the output window. It is reproduced below.

```
                                The SAS System                                1
                                          12:05 Wednesday, February 3, 1999

          Plot of TEMP*TEMPLAG.  Legend: A = 1 obs, B = 2 obs, etc.
```




```
NOTE: 1 obs had missing values.
```

SAS afficianadoes probably know how to use some tool that pops up a nice window (is it Insight?) but the SAS help system was not sufficiently informative that I could figure it out. Let me know if you know how to do it.

In summary, I found this exercise to be a royal pain in the SAS. I recommend using Splus.

3 Using S-Plus to do the work.

Now we look to a much nicer stat package for the kinds of things we had to work our SAS off to get done in the previous section. In Splus, just type in commands in the X-window at the Splus prompt. Invoke Splus in unix with "Splus" at the unix prompt (or "/usr/site/splus-3.4/bin/Splus" if you want to use the old version):

```
long-eared% /net/owlnet-d/splus-3.4/sunos5/Splus
License Warning : S-PLUS license expires Sun Feb 28 23:59:59 1999
S-PLUS : Copyright (c) 1988, 1996 MathSoft, Inc.
S : Copyright AT&T.
Version 3.4 Release 1 for Sun SPARC, SunOS 5.3 : 1996
Working data will be in /home/dcox/.Data
> #LINES BEGINNING WITH # ARE MY COMMENTS
> #getting a nice help window to pop up:
> help.start()
> #setting up my graphics window
> X11()
> #reading in the data
> cow_scan("cow.dat")
> #getting the time series plot
> tsplot(cow)
> title(main="Time Series plot of Cow Temp Data")
> #I clicked the print button in the graphics window
> #you may have to pull down menu to get it.
> #the lagged scatterplot is pretty easy:
> length(cow)
[1] 75
```

```
> plot(cow[1:74],cow[2:75])
> title(main="lagged scatterplot of Cow Temp")
> #getting the sample autocorrelation:
> cow.acf_acf(cow)
> #how nice - the plot automatically appeared in my graphics window
> #(I knew that was going to happen because I had a help window on acf).
> #reading off the values of the sample acf
> cow.acf
$acf:

, , 1

             [,1]
 [1,] 1.00000000
 [2,] 0.41050526
 [3,] 0.42295676
 [4,] 0.40500629
 [5,] 0.37264368
 [6,] 0.16299552
 [7,] 0.23204538
 [8,] 0.28196126
 [9,] 0.17605424
[10,] 0.03111671
[11,] 0.13753711
[12,] 0.05245340
[13,] 0.07643106
[14,] 0.09952299
[15,] 0.20836592
```

```
[16,] 0.20593612
[17,] 0.07179738
[18,] 0.07695194
[19,] 0.12802924

$lag:

, , 1

      [,1]
 [1,]    0
 [2,]    1
 [3,]    2
 [4,]    3
 [5,]    4
 [6,]    5
 [7,]    6
 [8,]    7
 [9,]    8
[10,]    9
[11,]   10
[12,]   11
[13,]   12
[14,]   13
[15,]   14
[16,]   15
[17,]   16
[18,]   17
```

```
[19,]   18

$n.used:
[1] 75

$type:
[1] "correlation"

$series:
[1] "cow"
> #more than I really wanted, but I see the acf at lags 1 and 2 are
> #0.41050526 and 0.42295676
> #that's all for now
> q()
```

When the session is over, there were three new files in the working directory: ps.out.0001.ps, ps.out.0002.ps, and ps.out.0003.ps. These are postscript files which are shown below.

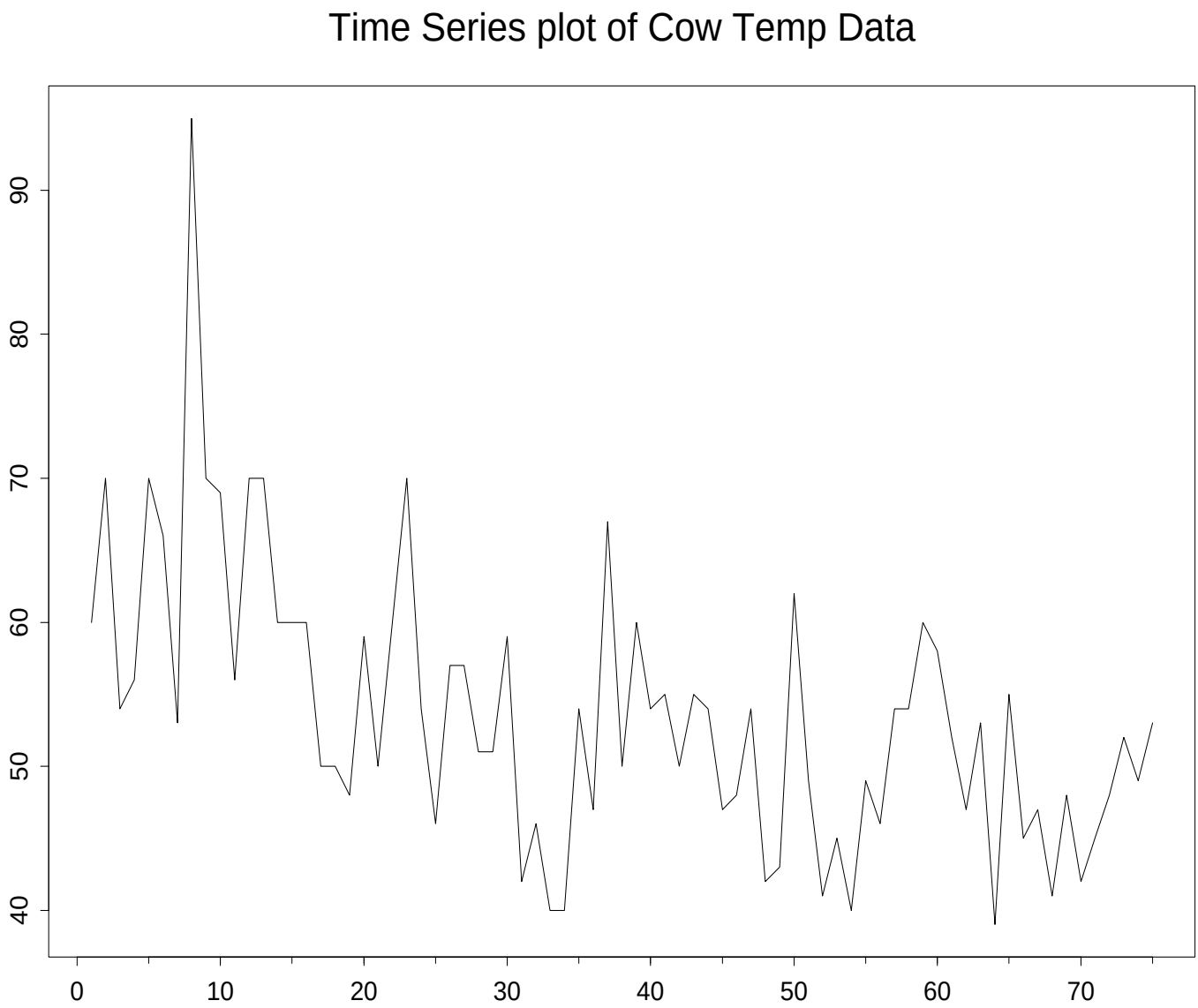


Figure 1: First plot.

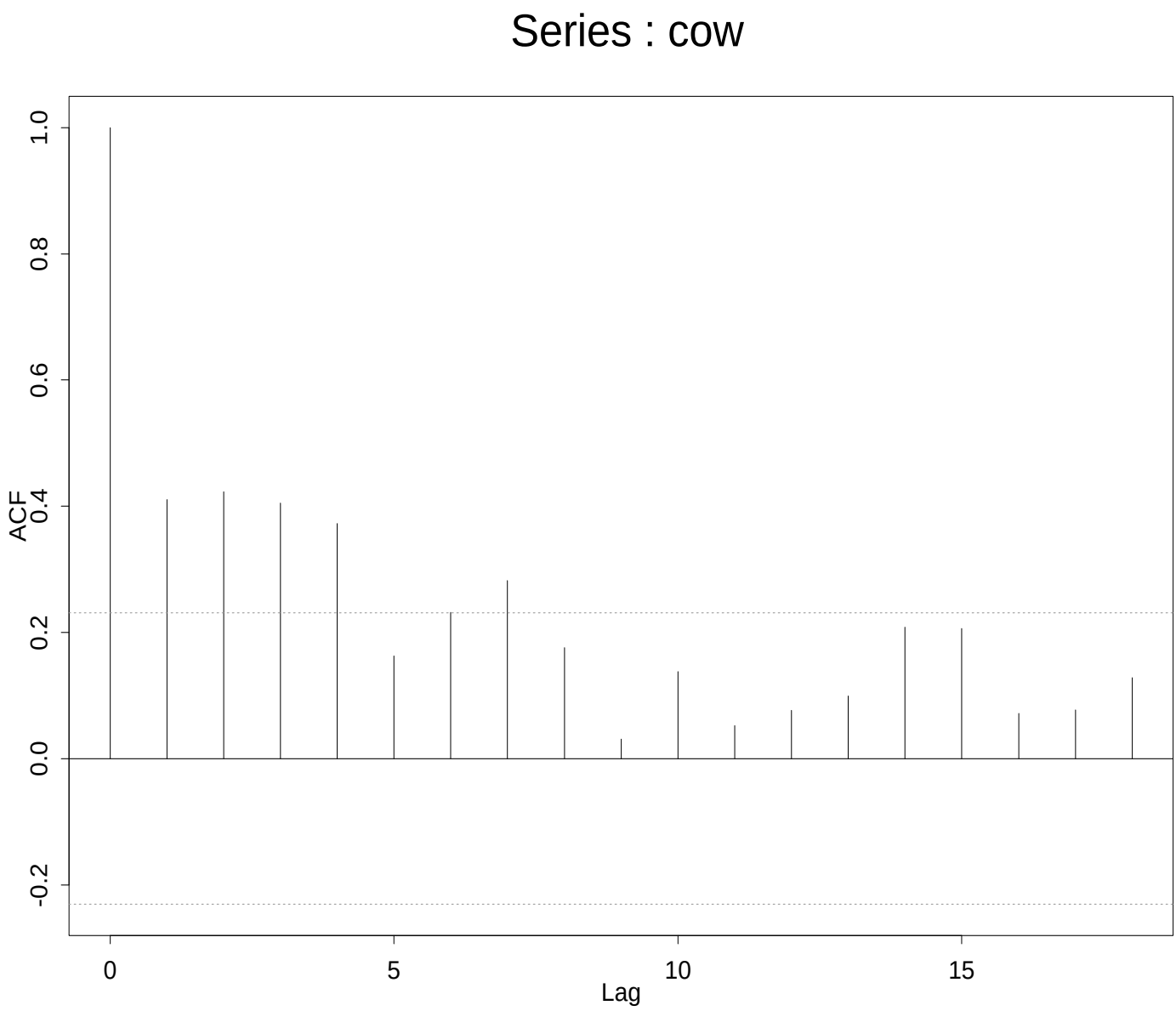


Figure 2: Second plot.

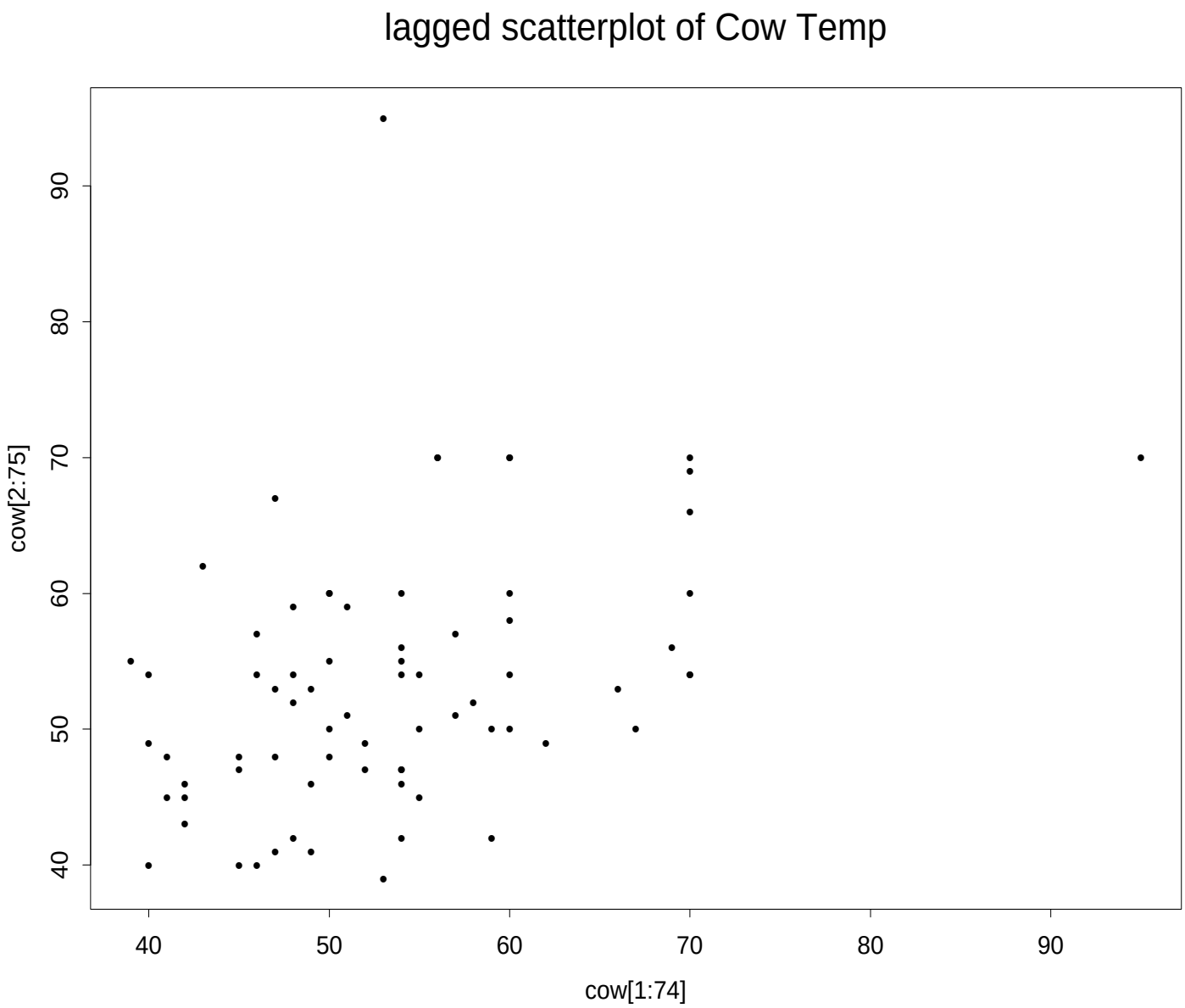


Figure 3: Third plot.