```
Your temporary usage period for IBM SPSS Statistics will expire in 3488 days.

* ============================================================.
* Box Plot Interpretation - CORRECTED SPSS Syntax Transcript.
*
* Topic:
* Box Plot Interpretation
*
* Existing cleaned input file:
* D:\DATA ANALYSIS\A Basic Descriptive Statistics Guides\spss_ready_data.csv
*
* Correct output folder created inside:
* D:\DATA ANALYSIS\A Basic Descriptive Statistics Guides\Box Plot Interpretati
on\SPSS
*
* This corrected transcript removes the earlier SPSS errors:
* 1. No mixed string/numeric COUNT command.
* 2. No invalid PERCENTILES keyword inside /STATISTICS.
* 3. No unsupported AGGREGATE PCT() function.
* 4. School and sex are converted into numeric grouping variables before EXAMI
NE.
* 5. G3 box-plot fences are added from verified Python/R summary:
*    Q1 = 10, Median = 12, Q3 = 14, IQR = 4, lower fence = 4, upper fence = 20
.
* ============================================================.

SET UNICODE=ON.
SET DECIMAL=DOT.
SET PRINTBACK=ON.
SET TNUMBERS=VALUES.
SET TVARS=LABELS.

* ------------------------------------------------------------.
* 1. Create topic and SPSS output folders inside the correct base folder.
* ------------------------------------------------------------.

HOST COMMAND=['cmd /c if not exist "D:\DATA ANALYSIS\A Basic Descriptive Stati
stics Guides\Box Plot Interpretation" mkdir "D:\DATA ANALYSIS\A Basic Descript
ive Statistics Guides\Box Plot Interpretation"'].
```

Host

```
[DataSet0]


HOST COMMAND=['cmd /c if not exist "D:\DATA ANALYSIS\A Basic Descriptive Stati
stics Guides\Box Plot Interpretation\SPSS" mkdir "D:\DATA ANALYSIS\A Basic Des
criptive Statistics Guides\Box Plot Interpretation\SPSS"'].
```

Host

```
FILE HANDLE basedir /NAME="D:\DATA ANALYSIS\A Basic Descriptive Statistics Gui
des".
FILE HANDLE topicdir /NAME="D:\DATA ANALYSIS\A Basic Descriptive Statistics Gu
ides\Box Plot Interpretation".
FILE HANDLE spssout /NAME="D:\DATA ANALYSIS\A Basic Descriptive Statistics Gui
des\Box Plot Interpretation\SPSS".

* ------------------------------------------------------------.
* 2. Import existing cleaned dataset.
* ------------------------------------------------------------.

GET DATA
 /TYPE=TXT
 /FILE="D:\DATA ANALYSIS\A Basic Descriptive Statistics Guides\spss_ready_data
.csv"
 /ENCODING='UTF8'
 /DELCASE=LINE
 /DELIMITERS=","
 /QUALIFIER='"'
 /ARRANGEMENT=DELIMITED
 /FIRSTCASE=2
 /IMPORTCASE=ALL
 /VARIABLES=
 school A20
 sex A20
 age F8.2
 address A20
 famsize A20
 Pstatus A20
 Medu F8.2
 Fedu F8.2
```

```
 Mjob A30
 Fjob A30
 reason A30
 guardian A30
 traveltime F8.2
 studytime F8.2
 failures F8.2
 schoolsup A20
 famsup A20
 paid A20
 activities A20
 nursery A20
 higher A20
 internet A20
 romantic A20
 famrel F8.2
 freetime F8.2
 goout F8.2
 Dalc F8.2
 Walc F8.2
 health F8.2
 absences F8.2
 G1 F8.2
 G2 F8.2
 G3 F8.2.
CACHE.
EXECUTE.

DATASET NAME BoxPlotInterpretationClean WINDOW=FRONT.

* ------------------------------------------------------------.
* 3. Add case ID and clear variable labels.
* ------------------------------------------------------------.

COMPUTE case_id = $CASENUM.
FORMATS case_id (F8.0).
EXECUTE.

VARIABLE LABELS
 case_id "Case ID"
 school "School"
 sex "Sex"
```

```
 age "Student age"
 studytime "Weekly study time category"
 failures "Number of past class failures"
 absences "Number of school absences"
 G1 "First period grade"
 G2 "Second period grade"
 G3 "Final grade".

VALUE LABELS studytime
 1 "Less than 2 hours"
 2 "2 to 5 hours"
 3 "5 to 10 hours"
 4 "More than 10 hours".

VALUE LABELS failures
 0 "0 failures"
 1 "1 failure"
 2 "2 failures"
 3 "3 failures".
EXECUTE.

* ------------------------------------------------------------.
* 4. Create numeric grouping variables for school and sex.
* This avoids EXAMINE warnings about long string grouping variables.
* ------------------------------------------------------------.

NUMERIC school_group sex_group (F1.0).

DO IF (RTRIM(school) = "GP").
 COMPUTE school_group = 1.
ELSE IF (RTRIM(school) = "MS").
 COMPUTE school_group = 2.
END IF.

DO IF (RTRIM(sex) = "F").
 COMPUTE sex_group = 1.
ELSE IF (RTRIM(sex) = "M").
 COMPUTE sex_group = 2.
END IF.

VARIABLE LABELS
 school_group "School group"
```

```
 sex_group "Sex group".

VALUE LABELS school_group
 1 "GP"
 2 "MS".

VALUE LABELS sex_group
 1 "Female"
 2 "Male".

EXECUTE.

* ------------------------------------------------------------.
* 5. Correct missing-data check.
* Do not use COUNT on mixed string and numeric variables.
* ------------------------------------------------------------.

COMPUTE numeric_missing_count = SUM(
 MISSING(G1),
 MISSING(G2),
 MISSING(G3),
 MISSING(absences),
 MISSING(studytime),
 MISSING(failures)
).

COMPUTE string_missing_count = SUM(
 MISSING(school),
 MISSING(sex)
).

COMPUTE complete_boxplot_case = (numeric_missing_count = 0 AND string_missing_
count = 0).

VARIABLE LABELS
 numeric_missing_count "Missing numeric values for box-plot variables"
 string_missing_count "Missing string values for box-plot grouping variables"
 complete_boxplot_case "Complete case for box-plot workflow".

VALUE LABELS complete_boxplot_case
 0 "Incomplete"
 1 "Complete".
```

```
EXECUTE.

FREQUENCIES VARIABLES=complete_boxplot_case numeric_missing_count string_missi
ng_count
 /ORDER=ANALYSIS.
```

Frequencies

```
[BoxPlotInterpretationClean]
```

Statistics

		Complete case for box-plot workflow	Missing numeric values for box-plot variables	Missing string values for box-plot grouping variables
N	Valid	649	649	649
	Missing	0	0	0

Frequency Table

Complete case for box-plot workflow

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	649	100.0	100.0	100.0

Missing numeric values for box-plot variables

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	649	100.0	100.0	100.0

Missing string values for box-plot grouping variables

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	649	100.0	100.0	100.0

```
* ------------------------------------------------------------.
* 6. Import and descriptive statistics check.
* ------------------------------------------------------------.

DESCRIPTIVES VARIABLES=G1 G2 G3 absences age studytime failures Medu Fedu
```

```
 /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
First period grade	649	.00	19.00	11.3991	2.74527
Second period grade	649	.00	19.00	11.5701	2.91364
Final grade	649	.00	19.00	11.9060	3.23066
Number of school absences	649	.00	32.00	3.6595	4.64076
Student age	649	15.00	22.00	16.7442	1.21814
Weekly study time category	649	1.00	4.00	1.9307	.82951
Number of past class failures	649	.00	3.00	.2219	.59324
Medu	649	.00	4.00	2.5146	1.13455
Fedu	649	.00	4.00	2.3066	1.09993
Valid N (listwise)	649				

```
FREQUENCIES VARIABLES=G1 G2 G3 absences
 /STATISTICS=MEAN MEDIAN STDDEV MINIMUM MAXIMUM SKEWNESS SESKEW KURTOSIS SEKUR
T
 /PERCENTILES=25 50 75
 /ORDER=ANALYSIS.
```

Frequencies

Statistics

		First period grade	Second period grade	Final grade	Number of school absences
N	Valid	649	649	649	649
	Missing	0	0	0	0
Mean		11.3991	11.5701	11.9060	3.6595
Median		11.0000	11.0000	12.0000	2.0000
Std. Deviation		2.74527	2.91364	3.23066	4.64076
Skewness		-.003	-.360	-.913	2.021
Std. Error of Skewness		.096	.096	.096	.096
Kurtosis		.037	1.662	2.712	5.781
Std. Error of Kurtosis		.192	.192	.192	.192
Minimum		.00	.00	.00	.00
Maximum		19.00	19.00	19.00	32.00
Percentiles	25	10.0000	10.0000	10.0000	.0000
	50	11.0000	11.0000	12.0000	2.0000
	75	13.0000	13.0000	14.0000	6.0000

Frequency Table

First period grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	1	.2	.2	.2
	4.00	2	.3	.3	.5
	5.00	5	.8	.8	1.2
	6.00	9	1.4	1.4	2.6
	7.00	33	5.1	5.1	7.7
	8.00	42	6.5	6.5	14.2
	9.00	65	10.0	10.0	24.2
	10.00	95	14.6	14.6	38.8
	11.00	91	14.0	14.0	52.9
	12.00	82	12.6	12.6	65.5
	13.00	72	11.1	11.1	76.6
	14.00	71	10.9	10.9	87.5
	15.00	35	5.4	5.4	92.9
	16.00	22	3.4	3.4	96.3
	17.00	16	2.5	2.5	98.8
	18.00	7	1.1	1.1	99.8
	19.00	1	.2	.2	100.0
	Total	649	100.0	100.0	

Second period grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	7	1.1	1.1	1.1
	5.00	3	.5	.5	1.5
	6.00	7	1.1	1.1	2.6
	7.00	16	2.5	2.5	5.1
	8.00	40	6.2	6.2	11.2
	9.00	72	11.1	11.1	22.3
	10.00	83	12.8	12.8	35.1
	11.00	103	15.9	15.9	51.0
	12.00	86	13.3	13.3	64.3
	13.00	80	12.3	12.3	76.6
	14.00	54	8.3	8.3	84.9
	15.00	38	5.9	5.9	90.8
	16.00	25	3.9	3.9	94.6
	17.00	20	3.1	3.1	97.7
	18.00	14	2.2	2.2	99.8
	19.00	1	.2	.2	100.0
	Total	649	100.0	100.0	

Final grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	15	2.3	2.3	2.3
	1.00	1	.2	.2	2.5
	5.00	1	.2	.2	2.6
	6.00	3	.5	.5	3.1
	7.00	10	1.5	1.5	4.6
	8.00	35	5.4	5.4	10.0
	9.00	35	5.4	5.4	15.4
	10.00	97	14.9	14.9	30.4
	11.00	104	16.0	16.0	46.4
	12.00	72	11.1	11.1	57.5
	13.00	82	12.6	12.6	70.1
	14.00	63	9.7	9.7	79.8
	15.00	49	7.6	7.6	87.4
	16.00	36	5.5	5.5	92.9
	17.00	29	4.5	4.5	97.4
	18.00	15	2.3	2.3	99.7
	19.00	2	.3	.3	100.0
	Total	649	100.0	100.0	

Number of school absences

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	244	37.6	37.6	37.6
	1.00	12	1.8	1.8	39.4
	2.00	110	16.9	16.9	56.4
	3.00	7	1.1	1.1	57.5
	4.00	93	14.3	14.3	71.8
	5.00	12	1.8	1.8	73.7
	6.00	49	7.6	7.6	81.2
	7.00	3	.5	.5	81.7
	8.00	42	6.5	6.5	88.1
	9.00	7	1.1	1.1	89.2
	10.00	21	3.2	3.2	92.4
	11.00	5	.8	.8	93.2
	12.00	12	1.8	1.8	95.1
	13.00	1	.2	.2	95.2
	14.00	8	1.2	1.2	96.5
	15.00	2	.3	.3	96.8
	16.00	10	1.5	1.5	98.3
	18.00	3	.5	.5	98.8
	21.00	2	.3	.3	99.1
	22.00	2	.3	.3	99.4
	24.00	1	.2	.2	99.5
	26.00	1	.2	.2	99.7
	30.00	1	.2	.2	99.8
	32.00	1	.2	.2	100.0
	Total	649	100.0	100.0	

```
* ------------------------------------------------------------.
* 7. Main EXAMINE box-plot statistics.
* Correct syntax: PERCENTILES is not placed under /STATISTICS.
* ------------------------------------------------------------.

EXAMINE VARIABLES=G1 G2 G3 absences
 /PLOT BOXPLOT STEMLEAF
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
```

```
/MISSING LISTWISE
/NOTOTAL.
```

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
First period grade	649	100.0%	0	0.0%	649	100.0%
Second period grade	649	100.0%	0	0.0%	649	100.0%
Final grade	649	100.0%	0	0.0%	649	100.0%
Number of school absences	649	100.0%	0	0.0%	649	100.0%

Descriptives

			Statistic	Std. Error
First period grade	Mean		11.3991	.10776
	95% Confidence Interval for Mean	Lower Bound	11.1875	
		Upper Bound	11.6107	
	5% Trimmed Mean		11.3861	
	Median		11.0000	
	Variance		7.536	
	Std. Deviation		2.74527	
	Minimum		.00	
	Maximum		19.00	
	Range		19.00	
	Interquartile Range		3.00	
	Skewness		-.003	.096
	Kurtosis		.037	.192
Second period grade	Mean		11.5701	.11437
	95% Confidence Interval for Mean	Lower Bound	11.3455	
		Upper Bound	11.7947	
	5% Trimmed Mean		11.6011	
	Median		11.0000	
	Variance		8.489	
	Std. Deviation		2.91364	
	Minimum		.00	

Descriptives

			Statistic	Std. Error
	Maximum		19.00	
	Range		19.00	
	Interquartile Range		3.00	
	Skewness		-.360	.096
	Kurtosis		1.662	.192
Final grade	Mean		11.9060	.12681
	95% Confidence Interval for Mean	Lower Bound	11.6570	
		Upper Bound	12.1550	
	5% Trimmed Mean		12.0574	
	Median		12.0000	
	Variance		10.437	
	Std. Deviation		3.23066	
	Minimum		.00	
	Maximum		19.00	
	Range		19.00	
	Interquartile Range		4.00	
	Skewness		-.913	.096
	Kurtosis		2.712	.192
Number of school absences	Mean		3.6595	.18217
	95% Confidence Interval for Mean	Lower Bound	3.3018	
		Upper Bound	4.0172	
	5% Trimmed Mean		3.0861	
	Median		2.0000	
	Variance		21.537	
	Std. Deviation		4.64076	
	Minimum		.00	
	Maximum		32.00	
	Range		32.00	
	Interquartile Range		6.00	
	Skewness		2.021	.096
	Kurtosis		5.781	.192

Extreme Values

			Case Number	Value
First period grade	Highest	1	618	19.00
		2	114	18.00
		3	333	18.00
		4	339	18.00
		5	345	18.00[a]
	Lowest	1	1	.00
		2	570	4.00
		3	568	4.00
		4	640	5.00
		5	606	5.00[b]
Second period grade	Highest	1	339	19.00
		2	197	18.00
		3	241	18.00
		4	333	18.00
		5	338	18.00[a]
	Lowest	1	611	.00
		2	606	.00
		3	604	.00
		4	598	.00
		5	568	.00[c]
Final grade	Highest	1	339	19.00
		2	637	19.00
		3	114	18.00
		4	182	18.00
		5	186	18.00[a]
	Lowest	1	641	.00
		2	640	.00
		3	638	.00
		4	627	.00
		5	611	.00[c]
Number of school absences	Highest	1	198	32.00
		2	213	30.00

Extreme Values

		Case Number	Value
	3	257	26.00
	4	151	24.00
	5	156	22.00[d]
Lowest	1	643	.00
	2	642	.00
	3	641	.00
	4	640	.00
	5	638	.00[c]

a. Only a partial list of cases with the value 18.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 5.00 are shown in the table of lower extremes.

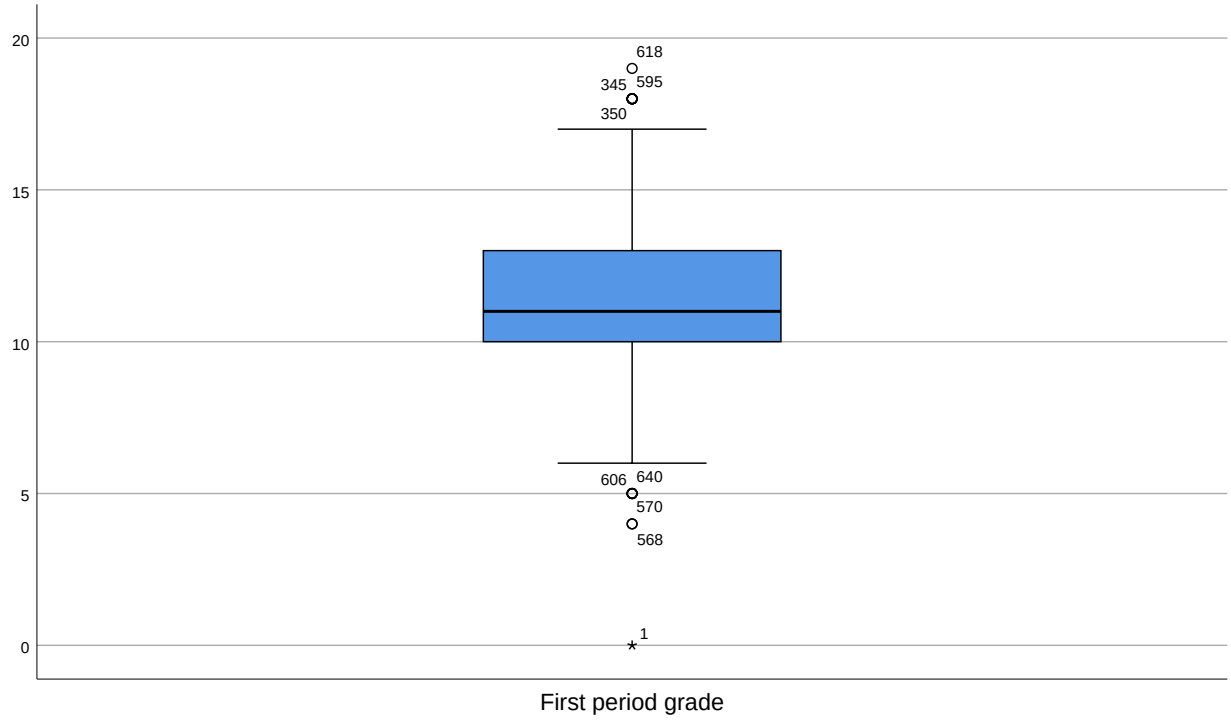
c. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

d. Only a partial list of cases with the value 22.00 are shown in the table of upper extremes.

First period grade

```
First period grade Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     8.00 Extremes    (=<5.0)
     9.00        6 .  000000000
    33.00        7 .  000000000000000000000000000000000
    42.00        8 .  000000000000000000000000000000000000000000
    65.00        9 .  00000000000000000000000000000000000000000000000000000000000000000
000000000
    95.00       10 .  00000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000
    91.00       11 .  00000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000
    82.00       12 .  00000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000
    72.00       13 .  00000000000000000000000000000000000000000000000000000000000000000
00000000000000000
```

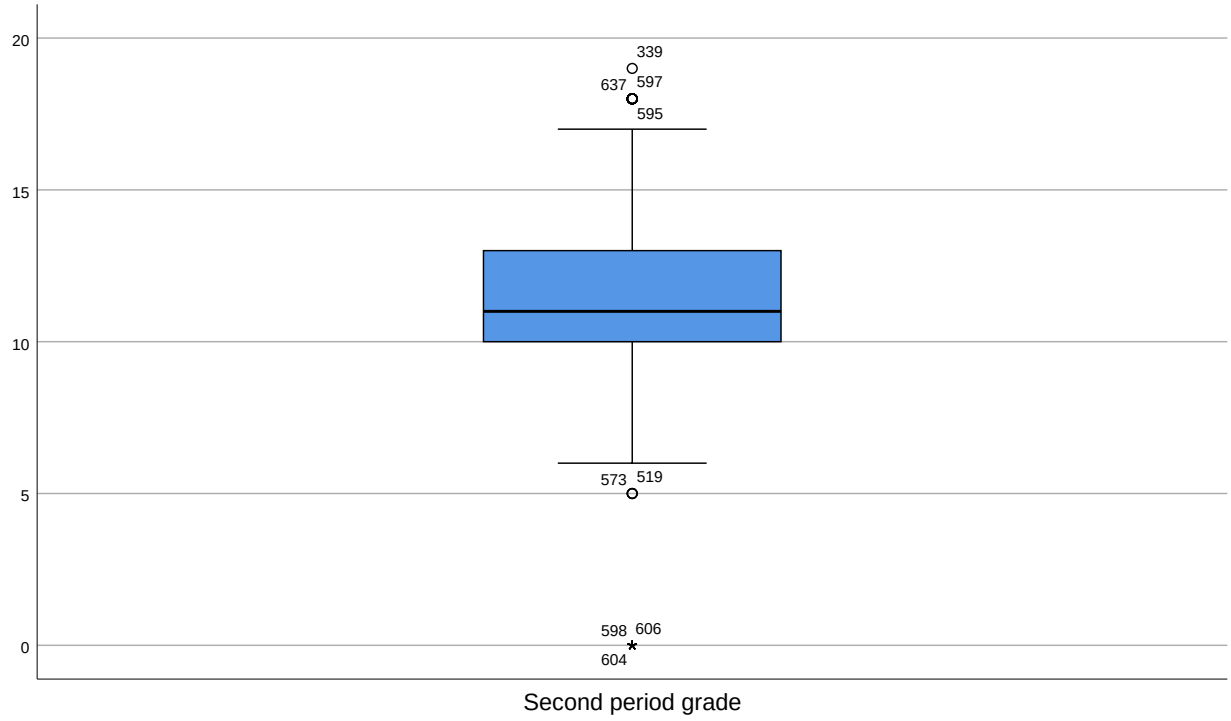
```
    71.00       14 .  00000000000000000000000000000000000000000000000000000000000000
000000000000000
    35.00       15 .  000000000000000000000000000000000
    22.00       16 .  0000000000000000000
    16.00       17 .  00000000000000000
     8.00 Extremes    (>=18.0)

 Stem width:      1.00
 Each leaf:        1 case(s)
```



Second period grade

```
Second period grade Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

    10.00 Extremes    (=<5.0)
     7.00        6 .  000
    16.00        7 .  00000000
    40.00        8 .  0000000000000000000
    72.00        9 .  000000000000000000000000000000000000
```

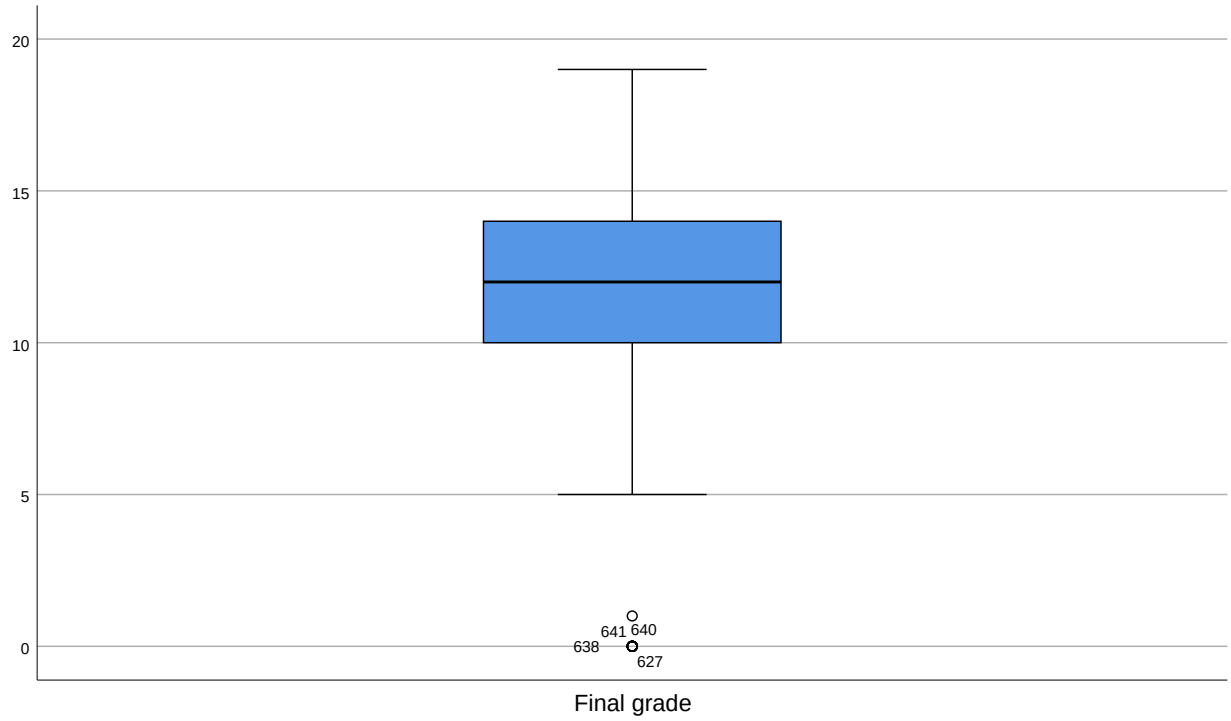
```
   83.00       10 .  00000000000000000000000000000000000000
  103.00       11 .  00000000000000000000000000000000000000000000000000
   86.00       12 .  00000000000000000000000000000000000000000000
   80.00       13 .  0000000000000000000000000000000000000000
   54.00       14 .  00000000000000000000000000
   38.00       15 .  0000000000000000000
   25.00       16 .  0000000000000
   20.00       17 .  0000000000
   15.00 Extremes    (>=18.0)

 Stem width:      1.00
 Each leaf:        2 case(s)
```



Final grade

```
Final grade Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

   16.00 Extremes    (=<1.0)
    1.00        5 .  &
```

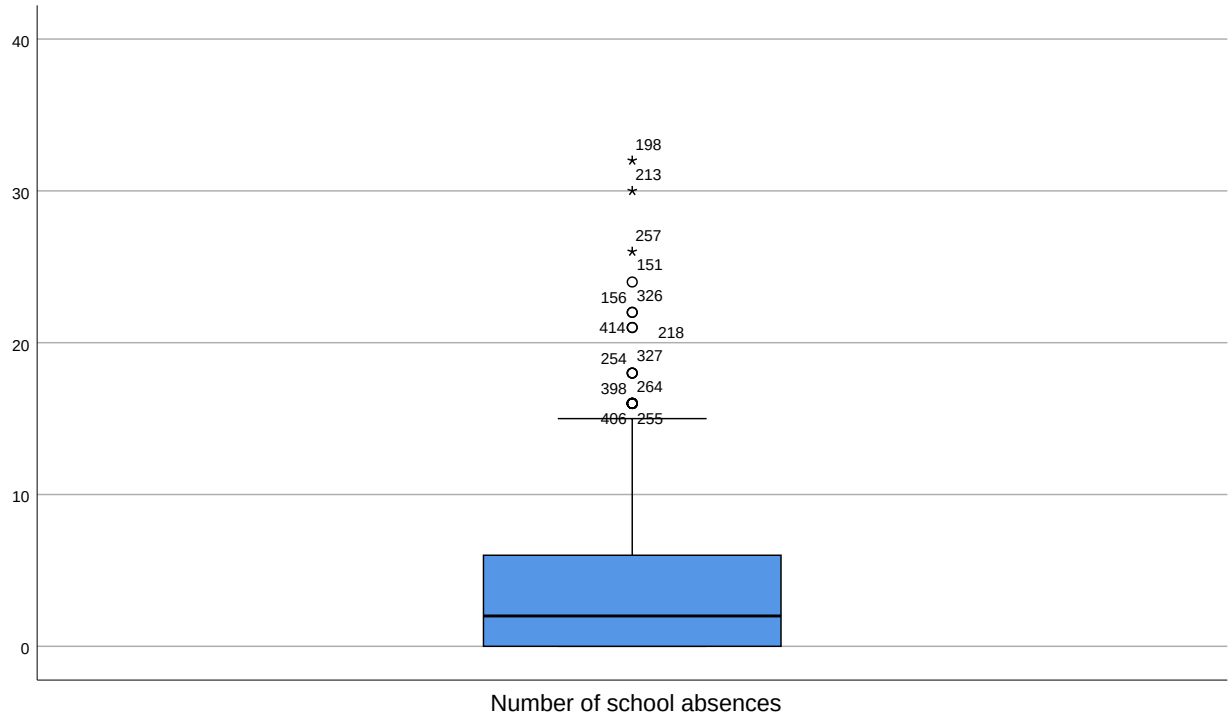
```
   3.00        6 .  0
  10.00        7 .  00000
  35.00        8 .  00000000000000000
  35.00        9 .  00000000000000000
  97.00       10 .  00000000000000000000000000000000000000000000000000
 104.00       11 .  000000000000000000000000000000000000000000000000000000
  72.00       12 .  000000000000000000000000000000000000
  82.00       13 .  0000000000000000000000000000000000000000
  63.00       14 .  00000000000000000000000000000000
  49.00       15 .  000000000000000000000000
  36.00       16 .  0000000000000000
  29.00       17 .  000000000000000
  15.00       18 .  0000000
   2.00       19 .  0

Stem width:      1.00
Each leaf:        2 case(s)


& denotes fractional leaves.
```



Number of school absences

Number of school absences Stem-and-Leaf Plot

```
 Frequency    Stem &  Leaf

   244.00        0 .  00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
    12.00        1 .  0000
   110.00        2 .  0000000000000000000000000000000000000
     7.00        3 .  00
    93.00        4 .  000000000000000000000000000000
    12.00        5 .  0000
    49.00        6 .  00000000000000000
     3.00        7 .  0
    42.00        8 .  000000000000000
     7.00        9 .  00
    21.00       10 .  0000000
     5.00       11 .  00
    12.00       12 .  0000
     1.00       13 .  &
     8.00       14 .  000
     2.00       15 .  0
    21.00 Extremes    (>=16.0)

 Stem width:      1.00
 Each leaf:        3 case(s)


 & denotes fractional leaves.
```




```
* ------------------------------------------------------------.
* 8. Grouped box plots for G3 using numeric grouping variables.
* ------------------------------------------------------------.

EXAMINE VARIABLES=G3 BY school_group
 /PLOT BOXPLOT
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

Explore

School group

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	School group	N	Percent	N	Percent	N	Percent
Final grade	1	423	100.0%	0	0.0%	423	100.0%
	2	226	100.0%	0	0.0%	226	100.0%

Descriptives

	School group			Statistic	Std. Error
Final grade	1	Mean		12.5768	.12766
		95% Confidence Interval for Mean	Lower Bound	12.3259	
			Upper Bound	12.8278	
		5% Trimmed Mean		12.6221	
		Median		13.0000	
		Variance		6.894	
		Std. Deviation		2.62564	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		3.00	
		Skewness		-.336	.119
		Kurtosis		1.388	.237
	2	Mean		10.6504	.25503
		95% Confidence Interval for Mean	Lower Bound	10.1479	
			Upper Bound	11.1530	
		5% Trimmed Mean		10.8500	
		Median		11.0000	
		Variance		14.699	
		Std. Deviation		3.83399	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		4.00	
		Skewness		-.828	.162
		Kurtosis		1.778	.322

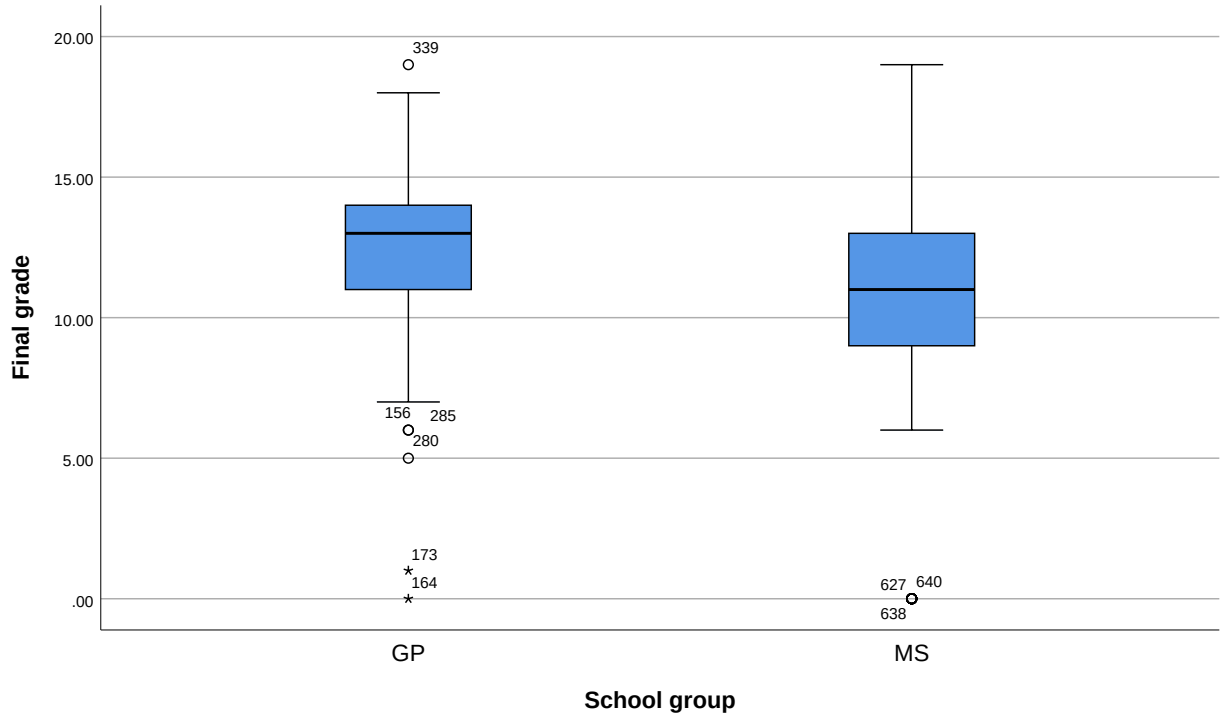
Extreme Values

	School group			Case Number	Value
Final grade	1	Highest	1	339	19.00
			2	114	18.00
			3	182	18.00
			4	186	18.00
			5	315	18.00[a]
		Lowest	1	164	.00
			2	173	1.00
			3	280	5.00
			4	285	6.00
			5	156	6.00
	2	Highest	1	637	19.00
			2	510	18.00
			3	550	18.00
			4	595	18.00
			5	597	18.00[a]
		Lowest	1	641	.00
			2	640	.00
			3	638	.00
			4	627	.00
			5	611	.00[b]

a. Only a partial list of cases with the value 18.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

Final grade




```
EXAMINE VARIABLES=G3 BY sex_group
 /PLOT BOXPLOT
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

Explore

Sex group

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Sex group	N	Percent	N	Percent	N	Percent
Final grade	1	383	100.0%	0	0.0%	383	100.0%
	2	266	100.0%	0	0.0%	266	100.0%

Descriptives

	Sex group			Statistic	Std. Error
Final grade	1	Mean		12.2533	.15964
		95% Confidence Interval for Mean	Lower Bound	11.9394	
			Upper Bound	12.5671	
		5% Trimmed Mean		12.3709	
		Median		12.0000	
		Variance		9.760	
		Std. Deviation		3.12415	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		4.00	
		Skewness		-.857	.125
		Kurtosis		2.683	.249
	2	Mean		11.4060	.20360
		95% Confidence Interval for Mean	Lower Bound	11.0051	
			Upper Bound	11.8069	
		5% Trimmed Mean		11.5961	
		Median		11.0000	
		Variance		11.027	
		Std. Deviation		3.32069	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		3.00	
		Skewness		-.980	.149
		Kurtosis		2.803	.298

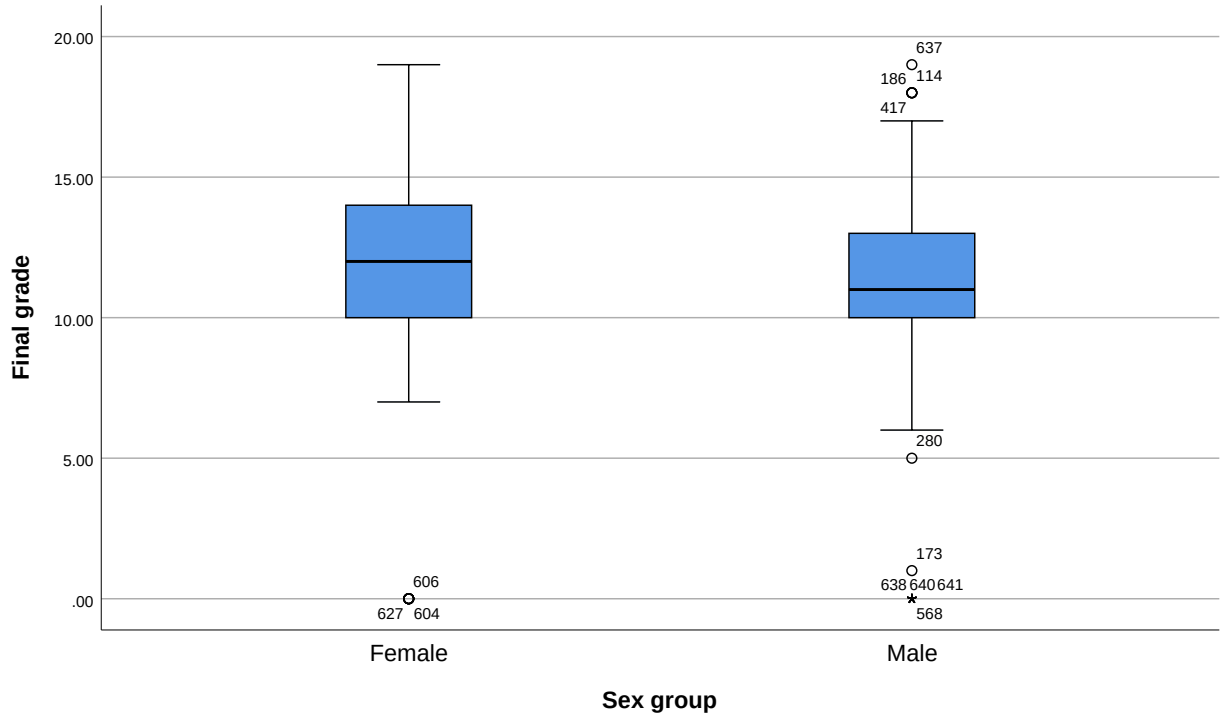
Extreme Values

	Sex group			Case Number	Value
Final grade	1	Highest	1	339	19.00
			2	182	18.00
			3	328	18.00
			4	333	18.00
			5	338	18.00[a]
		Lowest	1	627	.00
			2	611	.00
			3	606	.00
			4	604	.00
			5	598	.00[b]
	2	Highest	1	637	19.00
			2	114	18.00
			3	186	18.00
			4	315	18.00
			5	417	18.00
		Lowest	1	641	.00
			2	640	.00
			3	638	.00
			4	568	.00
			5	564	.00[b]

a. Only a partial list of cases with the value 18.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

Final grade




```
EXAMINE VARIABLES=G3 BY studytime
 /PLOT BOXPLOT
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

Explore

Weekly study time category

Case Processing Summary

		Cases				
		Valid		Missing		Total
	Weekly study time category	N	Percent	N	Percent	N
Final grade	1.00	212	100.0%	0	0.0%	212
	2.00	305	100.0%	0	0.0%	305
	3.00	97	100.0%	0	0.0%	97
	4.00	35	100.0%	0	0.0%	35

Case Processing Summary

		Cases	
		Total	
	Weekly study time category	Percent	
Final grade	1.00	100.0%	
	2.00	100.0%	
	3.00	100.0%	
	4.00	100.0%	

Descriptives

	Weekly study time category			Statistic	Std. Error
Final grade	1.00	Mean		10.8443	.22106
		95% Confidence Interval for Mean	Lower Bound	10.4086	
			Upper Bound	11.2801	
		5% Trimmed Mean		11.0419	
		Median		11.0000	
		Variance		10.360	
		Std. Deviation		3.21862	
		Minimum		.00	
		Maximum		18.00	
		Range		18.00	
		Interquartile Range		3.00	
		Skewness		-1.078	.167
		Kurtosis		3.117	.333
	2.00	Mean		12.0918	.18570
		95% Confidence Interval for Mean	Lower Bound	11.7264	
			Upper Bound	12.4572	
		5% Trimmed Mean		12.2505	
		Median		12.0000	
		Variance		10.518	
		Std. Deviation		3.24313	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		4.00	
		Skewness		-1.028	.140
		Kurtosis		3.044	.278

Descriptives

Weekly study time category			Statistic	Std. Error
3.00	Mean		13.2268	.25405
	95% Confidence Interval for Mean	Lower Bound	12.7225	
		Upper Bound	13.7311	
	5% Trimmed Mean		13.2732	
	Median		13.0000	
	Variance		6.261	
	Std. Deviation		2.50210	
	Minimum		8.00	
	Maximum		18.00	
	Range		10.00	
	Interquartile Range		3.50	
	Skewness		-.190	.245
	Kurtosis		-.502	.485
4.00	Mean		13.0571	.51358
	95% Confidence Interval for Mean	Lower Bound	12.0134	
		Upper Bound	14.1009	
	5% Trimmed Mean		13.0714	
	Median		13.0000	
	Variance		9.232	
	Std. Deviation		3.03841	
	Minimum		6.00	
	Maximum		19.00	
	Range		13.00	
	Interquartile Range		4.00	
	Skewness		.209	.398
	Kurtosis		-.339	.778

Extreme Values

	Weekly study time category			Case Number	Value
Final grade	1.00	Highest	1	114	18.00
			2	16	17.00
			3	345	17.00
			4	357	17.00
			5	365	17.00[a]
		Lowest	1	641	.00
			2	640	.00
			3	638	.00
			4	606	.00
			5	584	.00[b]
	2.00	Highest	1	637	19.00
			2	182	18.00
			3	186	18.00
			4	315	18.00
			5	417	18.00[c]
		Lowest	1	627	.00
			2	611	.00
			3	604	.00
			4	598	.00
			5	587	.00[b]
	3.00	Highest	1	333	18.00
			2	338	18.00
			3	618	18.00
			4	358	17.00
			5	360	17.00[a]
		Lowest	1	586	8.00
			2	455	8.00
			3	454	8.00
			4	348	8.00
			5	220	8.00
	4.00	Highest	1	339	19.00
			2	328	18.00

Extreme Values

Weekly study time category		Case Number	Value
	3	350	18.00
	4	607	18.00
	5	48	17.00[a]
Lowest	1	524	6.00
	2	503	9.00
	3	498	10.00
	4	107	10.00
	5	106	10.00[d]

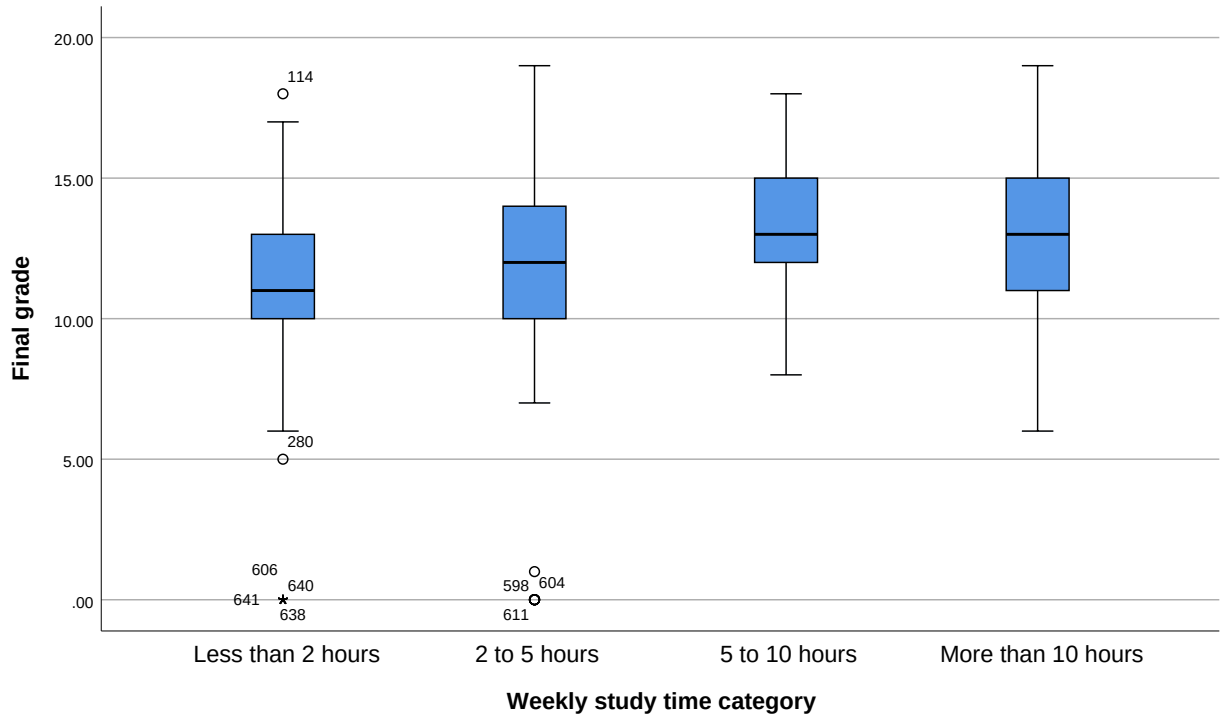
a. Only a partial list of cases with the value 17.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 18.00 are shown in the table of upper extremes.

d. Only a partial list of cases with the value 10.00 are shown in the table of lower extremes.

Final grade



```
EXAMINE VARIABLES=G3 BY failures
 /PLOT BOXPLOT
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

Explore

Number of past class failures

Case Processing Summary

		Cases				
		Valid		Missing		Total
	Number of past class failures	N	Percent	N	Percent	N
Final grade	.00	549	100.0%	0	0.0%	549
	1.00	70	100.0%	0	0.0%	70
	2.00	16	100.0%	0	0.0%	16
	3.00	14	100.0%	0	0.0%	14

Case Processing Summary

		Cases
	Number of past class failures	Total Percent
Final grade	.00	100.0%
	1.00	100.0%
	2.00	100.0%
	3.00	100.0%

Descriptives

	Number of past class failures			Statistic	Std. Error
Final grade	.00	Mean		12.5100	.12073
		95% Confidence Interval for Mean	Lower Bound	12.2729	
			Upper Bound	12.7472	
		5% Trimmed Mean		12.5840	
		Median		12.0000	
		Variance		8.002	
		Std. Deviation		2.82881	
		Minimum		.00	
		Maximum		19.00	
		Range		19.00	
		Interquartile Range		3.00	
		Skewness		-.696	.104
		Kurtosis		2.838	.208
	1.00	Mean		8.6429	.41155
		95% Confidence Interval for Mean	Lower Bound	7.8218	
			Upper Bound	9.4639	
		5% Trimmed Mean		8.8333	
		Median		10.0000	
		Variance		11.856	
		Std. Deviation		3.44327	
		Minimum		.00	
		Maximum		16.00	
		Range		16.00	
		Interquartile Range		2.00	
		Skewness		-1.503	.287
		Kurtosis		2.154	.566
	2.00	Mean		8.8125	.80218
		95% Confidence Interval for Mean	Lower Bound	7.1027	
			Upper Bound	10.5223	
		5% Trimmed Mean		8.9583	
		Median		9.5000	
		Variance		10.296	
		Std. Deviation		3.20871	
		Minimum		.00	
		Maximum		15.00	

Descriptives

Number of past class failures			Statistic	Std. Error
	Range		15.00	
	Interquartile Range		2.75	
	Skewness		-.996	.564
	Kurtosis		3.546	1.091
3.00	Mean		8.0714	.74468
	95% Confidence Interval for Mean	Lower Bound	6.4626	
		Upper Bound	9.6802	
	5% Trimmed Mean		8.3571	
	Median		8.5000	
	Variance		7.764	
	Std. Deviation		2.78635	
	Minimum		.00	
	Maximum		11.00	
	Range		11.00	
	Interquartile Range		2.25	
	Skewness		-2.077	.597
	Kurtosis		5.224	1.154

Extreme Values

	Number of past class failures			Case Number	Value
Final grade	.00	Highest	1	339	19.00
			2	637	19.00
			3	114	18.00
			4	182	18.00
			5	186	18.00[a]
		Lowest	1	638	.00
			2	627	.00
			3	604	.00
			4	587	.00
			5	520	.00[b]
	1.00	Highest	1	213	16.00
			2	321	13.00
			3	407	13.00
			4	560	13.00
			5	472	12.00
		Lowest	1	641	.00
			2	640	.00
			3	606	.00
			4	598	.00
			5	584	.00[b]
	2.00	Highest	1	352	15.00
			2	408	12.00
			3	128	11.00
			4	147	10.00
			5	174	10.00[c]
		Lowest	1	164	.00
			2	285	6.00
			3	591	7.00
			4	573	7.00
			5	582	8.00[d]
	3.00	Highest	1	171	11.00
			2	79	10.00

Extreme Values

Number of past class failures		Case Number	Value
3		170	10.00
4		238	10.00
5		558	10.00
Lowest	1	611	.00
	2	280	5.00
	3	19	7.00
	4	544	8.00
	5	479	8.00[d]

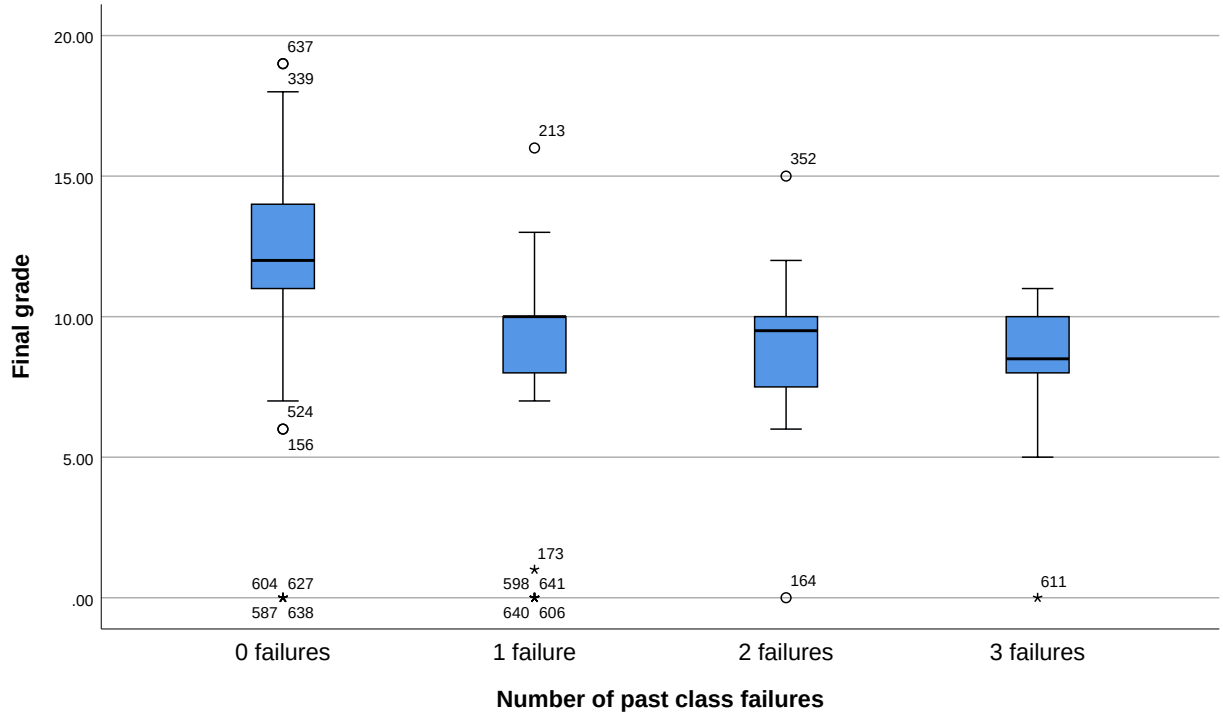
a. Only a partial list of cases with the value 18.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 10.00 are shown in the table of upper extremes.

d. Only a partial list of cases with the value 8.00 are shown in the table of lower extremes.

Final grade



```
EXAMINE VARIABLES=absences BY school_group
 /PLOT BOXPLOT
 /COMPARE GROUPS
 /STATISTICS DESCRIPTIVES EXTREME
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```

Explore

School group

Case Processing Summary

		Cases				
		Valid		Missing		Total
	School group	N	Percent	N	Percent	N
Number of school absences	1	423	100.0%	0	0.0%	423
	2	226	100.0%	0	0.0%	226

Case Processing Summary

		Cases
		Total
	School group	Percent
Number of school absences	1	100.0%
	2	100.0%

Descriptives

	School group			Statistic	Std. Error
Number of school absences	1	Mean		4.2151	.25244
		95% Confidence Interval for Mean	Lower Bound	3.7189	
			Upper Bound	4.7113	
		5% Trimmed Mean		3.5954	
		Median		2.0000	
		Variance		26.956	
		Std. Deviation		5.19191	
		Minimum		.00	
		Maximum		32.00	
		Range		32.00	
		Interquartile Range		6.00	
		Skewness		1.886	.119
		Kurtosis		4.573	.237
	2	Mean		2.6195	.20825
		95% Confidence Interval for Mean	Lower Bound	2.2091	
			Upper Bound	3.0298	
		5% Trimmed Mean		2.3220	
		Median		2.0000	
		Variance		9.801	
		Std. Deviation		3.13069	
		Minimum		.00	
		Maximum		12.00	
		Range		12.00	
		Interquartile Range		4.00	
		Skewness		1.125	.162
		Kurtosis		.428	.322

Extreme Values

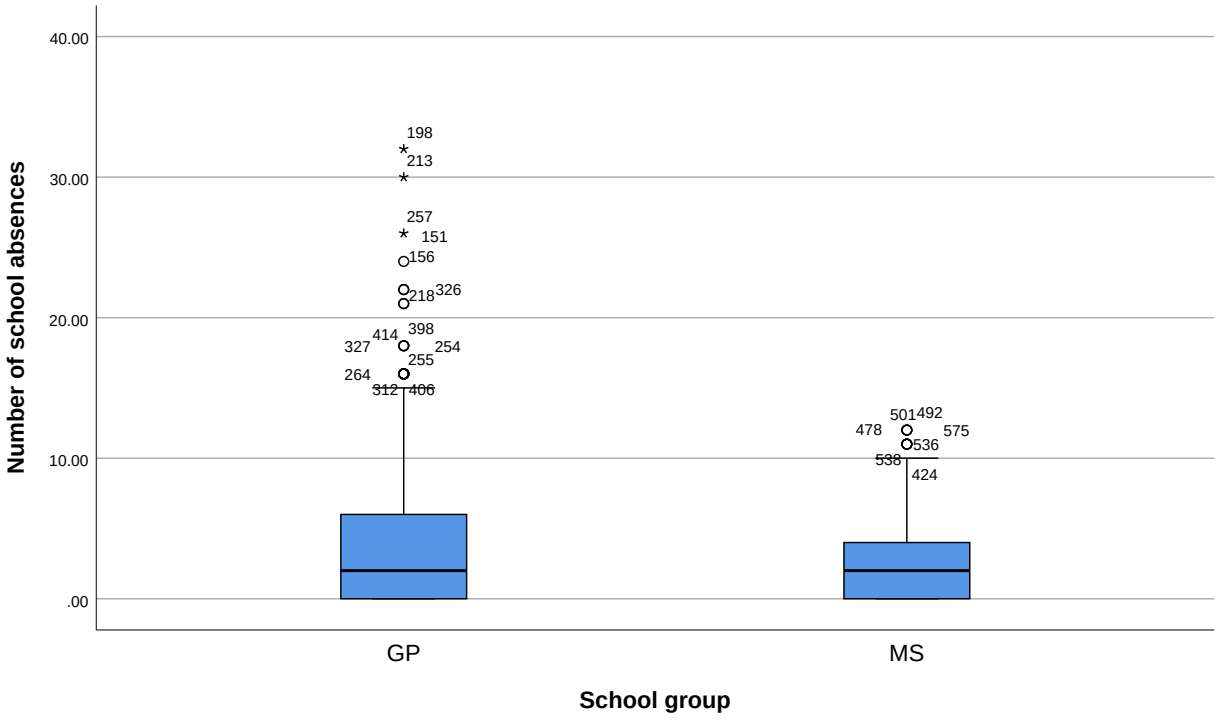
	School group			Case Number	Value
Number of school absences	1	Highest	1	198	32.00
			2	213	30.00
			3	257	26.00
			4	151	24.00
			5	156	22.00[a]
		Lowest	1	413	.00
			2	411	.00
			3	408	.00
			4	403	.00
			5	402	.00[b]
	2	Highest	1	492	12.00
			2	501	12.00
			3	575	12.00
			4	424	11.00
			5	478	11.00[c]
		Lowest	1	643	.00
			2	642	.00
			3	641	.00
			4	640	.00
			5	638	.00[b]

a. Only a partial list of cases with the value 22.00 are shown in the table of upper extremes.

b. Only a partial list of cases with the value .00 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 11.00 are shown in the table of upper extremes.

Number of school absences




```
* ------------------------------------------------------------.
* 9. Correct G3 box-plot fence and outlier-candidate calculation.
*
* Verified summary from Python/R chart panel:
* n = 649
* Q1 = 10
* Median = 12
* Q3 = 14
* IQR = 4
* Lower fence = 4
* Upper fence = 20
* Rule: below Q1 - 1.5*IQR or above Q3 + 1.5*IQR.
* ------------------------------------------------------------.

COMPUTE g3_q1 = 10.
COMPUTE g3_median = 12.
COMPUTE g3_q3 = 14.
COMPUTE g3_iqr = g3_q3 - g3_q1.
COMPUTE g3_lower_fence = g3_q1 - (1.5 * g3_iqr).
COMPUTE g3_upper_fence = g3_q3 + (1.5 * g3_iqr).
COMPUTE g3_boxplot_outlier = (G3 < g3_lower_fence OR G3 > g3_upper_fence).

VARIABLE LABELS
```

```
 g3_q1 "G3 first quartile"
 g3_median "G3 median"
 g3_q3 "G3 third quartile"
 g3_iqr "G3 interquartile range"
 g3_lower_fence "G3 lower box-plot fence"
 g3_upper_fence "G3 upper box-plot fence"
 g3_boxplot_outlier "G3 box-plot outlier candidate".

VALUE LABELS g3_boxplot_outlier
 0 "Not box-plot outlier"
 1 "Box-plot outlier candidate".

FORMATS g3_q1 g3_median g3_q3 g3_iqr g3_lower_fence g3_upper_fence (F8.2).
EXECUTE.

AGGREGATE
 /OUTFILE=* MODE=ADDVARIABLES
 /BREAK=
 /g3_valid_n=N(G3)
 /g3_outlier_count=SUM(g3_boxplot_outlier).

COMPUTE g3_outlier_percent = (g3_outlier_count / g3_valid_n) * 100.
FORMATS g3_outlier_percent (F8.2).
VARIABLE LABELS
 g3_valid_n "Valid N for G3"
 g3_outlier_count "Number of G3 box-plot outlier candidates"
 g3_outlier_percent "Percent of G3 box-plot outlier candidates".
EXECUTE.

FREQUENCIES VARIABLES=g3_boxplot_outlier
 /ORDER=ANALYSIS.
```

Frequencies

Statistics

G3 box-plot outlier candidate

N	Valid	649
	Missing	0

G3 box-plot outlier candidate

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	633	97.5	97.5	97.5
	1.00	16	2.5	2.5	100.0
	Total	649	100.0	100.0	

```
DESCRIPTIVES VARIABLES=g3_valid_n g3_q1 g3_median g3_q3 g3_iqr g3_lower_fence
g3_upper_fence g3_outlier_count g3_outlier_percent
 /STATISTICS=MEAN MIN MAX.
```

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean
Valid N for G3	649	649	649	649.00
G3 first quartile	649	10.00	10.00	10.0000
G3 median	649	12.00	12.00	12.0000
G3 third quartile	649	14.00	14.00	14.0000
G3 interquartile range	649	4.00	4.00	4.0000
G3 lower box-plot fence	649	4.00	4.00	4.0000
G3 upper box-plot fence	649	20.00	20.00	20.0000
Number of G3 box-plot outlier candidates	649	16.00	16.00	16.0000
Percent of G3 box-plot outlier candidates	649	2.47	2.47	2.4653
Valid N (listwise)	649			

```
SORT CASES BY g3_boxplot_outlier(D) G3(A) case_id(A).

LIST VARIABLES=case_id school sex studytime failures absences G1 G2 G3 g3_lowe
r_fence g3_upper_fence g3_boxplot_outlier
 /CASES=FROM 1 TO 50.
```

List

case_id	school	sex	g3_lower_fence	g3_upper_fence	g3_boxplot_outlier	studytime	failures	absences	G1	G2	G3
164	GP	M	4.00	20.00	1.00	1.00	2.00	.00	11.00	9.00	.00
441	MS	M	4.00	20.00	1.00	2.00	.00	.00	7.00	.00	.00
520	MS	M	4.00	20.00	1.00	2.00	.00	.00	8.00	7.00	.00
564	MS	M	4.00	20.00	1.00	1.00	1.00	.00	7.00	.00	.00
568	MS	M	4.00	20.00	1.00	1.00	1.00	.00	4.00	.00	.00
584	MS	F	4.00	20.00	1.00	1.00	1.00	.00	8.00	6.00	.00
587	MS	F	4.00	20.00	1.00	2.00	.00	.00	8.00	8.00	.00
598	MS	F	4.00	20.00	1.00	2.00	1.00	.00	9.00	.00	.00
604	MS	F	4.00	20.00	1.00	2.00	.00	.00	5.00	.00	.00
606	MS	F	4.00	20.00	1.00	1.00	1.00	.00	5.00	.00	.00
611	MS	F	4.00	20.00	1.00	2.00	3.00	.00	8.00	.00	.00
627	MS	F	4.00	20.00	1.00	2.00	.00	.00	7.00	5.00	.00
638	MS	M	4.00	20.00	1.00	1.00	.00	.00	7.00	7.00	.00
640	MS	M	4.00	20.00	1.00	1.00	1.00	.00	5.00	8.00	.00
641	MS	M	4.00	20.00	1.00	1.00	1.00	.00	7.00	7.00	.00
173	GP	M	4.00	20.00	1.00	2.00	1.00	.00	10.00	10.00	1.00
280	GP	M	4.00	20.00	.00	1.00	3.00	12.00	7.00	8.00	5.00

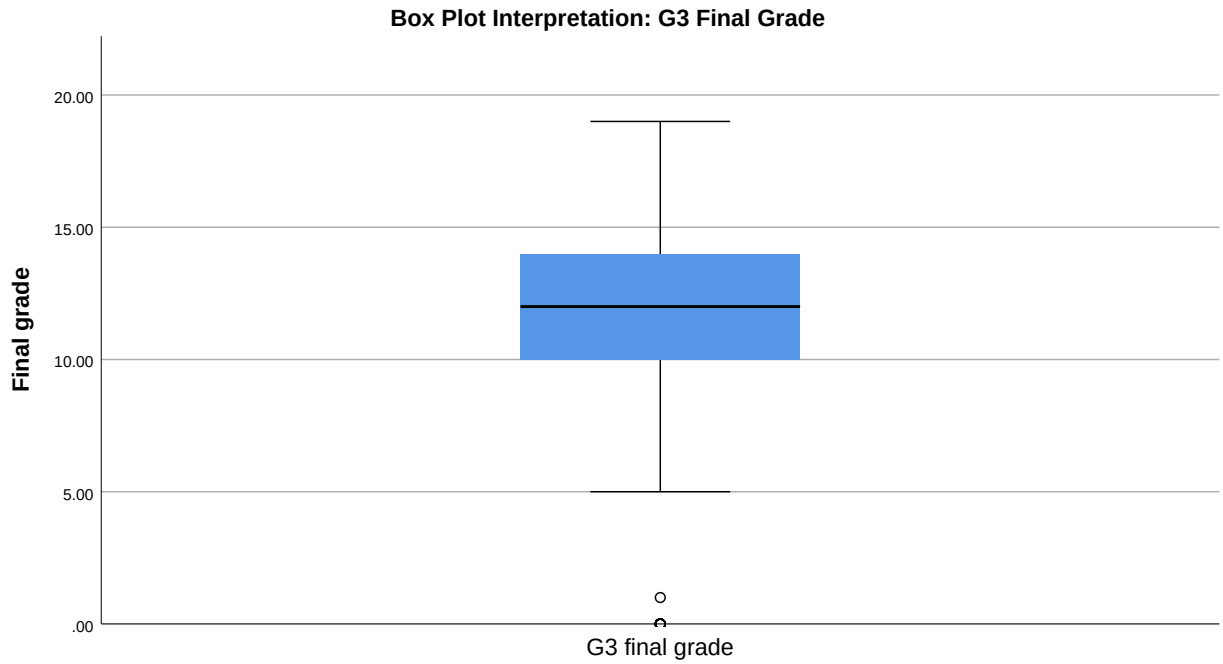
156	GP			M			1.00	.00	22.00
9.00		7.00	6.00	4.00	20.00	.00			
285	GP			M			1.00	2.00	4.00
7.00		8.00	6.00	4.00	20.00	.00			
524	MS			M			4.00	.00	2.00
5.00		6.00	6.00	4.00	20.00	.00			
19	GP			M			1.00	3.00	2.00
8.00		8.00	7.00	4.00	20.00	.00			
256	GP			F			1.00	1.00	14.00
8.00		7.00	7.00	4.00	20.00	.00			
283	GP			M			1.00	.00	8.00
7.00		8.00	7.00	4.00	20.00	.00			
433	MS			F			1.00	1.00	.00
6.00		6.00	7.00	4.00	20.00	.00			
501	MS			M			1.00	1.00	12.00
6.00		7.00	7.00	4.00	20.00	.00			
513	MS			F			2.00	1.00	9.00
7.00		7.00	7.00	4.00	20.00	.00			
515	MS			F			1.00	.00	2.00
7.00		6.00	7.00	4.00	20.00	.00			
573	MS			F			2.00	2.00	8.00
5.00		5.00	7.00	4.00	20.00	.00			
583	MS			F			1.00	.00	.00
6.00		8.00	7.00	4.00	20.00	.00			
591	MS			F			2.00	2.00	3.00
7.00		8.00	7.00	4.00	20.00	.00			
101	GP			M			1.00	.00	12.00
9.00		9.00	8.00	4.00	20.00	.00			
132	GP			F			2.00	3.00	10.00
10.00		9.00	8.00	4.00	20.00	.00			
162	GP			M			1.00	.00	16.00
9.00		9.00	8.00	4.00	20.00	.00			
169	GP			M			2.00	.00	.00
7.00		8.00	8.00	4.00	20.00	.00			
176	GP			M			2.00	2.00	4.00
7.00		6.00	8.00	4.00	20.00	.00			
178	GP			M			1.00	1.00	2.00
9.00		8.00	8.00	4.00	20.00	.00			
179	GP			M			2.00	.00	.00
8.00		8.00	8.00	4.00	20.00	.00			
180	GP			M			1.00	3.00	10.00
8.00		7.00	8.00	4.00	20.00	.00			

```
     220 GP                   F                        3.00     1.00    14.00
    8.00     9.00     8.00     4.00    20.00       .00
     240 GP                   F                        2.00      .00     4.00
    8.00     8.00     8.00     4.00    20.00       .00
     257 GP                   M                        1.00     1.00    26.00
    7.00     8.00     8.00     4.00    20.00       .00
     263 GP                   F                        1.00     1.00    14.00
    9.00     9.00     8.00     4.00    20.00       .00
     264 GP                   M                        2.00      .00    16.00
    8.00     7.00     8.00     4.00    20.00       .00
     348 GP                   F                        3.00      .00      .00
    7.00     7.00     8.00     4.00    20.00       .00
     437 MS                   F                        1.00     1.00     4.00
    6.00     7.00     8.00     4.00    20.00       .00
     444 MS                   M                        2.00      .00     7.00
    7.00     9.00     8.00     4.00    20.00       .00
     454 MS                   F                        3.00     1.00      .00
   10.00     8.00     8.00     4.00    20.00       .00
     455 MS                   F                        3.00     1.00     3.00
    9.00     8.00     8.00     4.00    20.00       .00
     466 MS                   F                        2.00     1.00     9.00
    9.00     8.00     8.00     4.00    20.00       .00
     477 MS                   F                        1.00      .00     2.00
   10.00     8.00     8.00     4.00    20.00       .00


Number of cases read:  50    Number of cases listed:  50

* ------------------------------------------------------------.
* 10. Save corrected case-level CSV with fence and outlier columns.
* ------------------------------------------------------------.

SAVE TRANSLATE OUTFILE="D:\DATA ANALYSIS\A Basic Descriptive Statistics Guides
\Box Plot Interpretation\SPSS\box_plot_interpretation_spss_case_level_data_COR
RECTED.csv"
 /TYPE=CSV
 /ENCODING='UTF8'
 /MAP
 /REPLACE
 /FIELDNAMES
 /CELLS=VALUES.
```

```
Data written to D:\DATA ANALYSIS\...\Box Plot Interpretation\SPSS\box_plot_int
erpretation_spss_case_level_data_CORRECTED.csv.
49 variables and 649 cases written.
Variable: school            Type: String   Width:  20
Variable: sex               Type: String   Width:  20
Variable: age               Type: Number   Width:   8   Dec: 2
Variable: address           Type: String   Width:  20
Variable: famsize           Type: String   Width:  20
Variable: Pstatus           Type: String   Width:  20
Variable: Medu              Type: Number   Width:   8   Dec: 2
Variable: Fedu              Type: Number   Width:   8   Dec: 2
Variable: Mjob              Type: String   Width:  30
Variable: Fjob              Type: String   Width:  30
Variable: reason            Type: String   Width:  30
Variable: guardian          Type: String   Width:  30
Variable: traveltime        Type: Number   Width:   8   Dec: 2
Variable: studytime         Type: Number   Width:   8   Dec: 2
Variable: failures          Type: Number   Width:   8   Dec: 2
Variable: schoolsup         Type: String   Width:  20
Variable: famsup            Type: String   Width:  20
Variable: paid              Type: String   Width:  20
Variable: activities        Type: String   Width:  20
Variable: nursery           Type: String   Width:  20
Variable: higher            Type: String   Width:  20
Variable: internet          Type: String   Width:  20
Variable: romantic          Type: String   Width:  20
Variable: famrel            Type: Number   Width:   8   Dec: 2
Variable: freetime          Type: Number   Width:   8   Dec: 2
Variable: goout             Type: Number   Width:   8   Dec: 2
Variable: Dalc              Type: Number   Width:   8   Dec: 2
Variable: Walc              Type: Number   Width:   8   Dec: 2
Variable: health            Type: Number   Width:   8   Dec: 2
Variable: absences          Type: Number   Width:   8   Dec: 2
Variable: G1                Type: Number   Width:   8   Dec: 2
Variable: G2                Type: Number   Width:   8   Dec: 2
Variable: G3                Type: Number   Width:   8   Dec: 2
Variable: case_id           Type: Number   Width:   8   Dec: 0
Variable: school_group      Type: Number   Width:   1   Dec: 0
Variable: sex_group         Type: Number   Width:   1   Dec: 0
Variable: numeric_missing_count   Type: Number   Width:   8   Dec: 2
Variable: string_missing_count   Type: Number   Width:   8   Dec: 2
Variable: complete_boxplot_case   Type: Number   Width:   8   Dec: 2
```

```
Variable: g3_q1              Type: Number   Width:   8   Dec: 2
Variable: g3_median          Type: Number   Width:   8   Dec: 2
Variable: g3_q3              Type: Number   Width:   8   Dec: 2
Variable: g3_iqr             Type: Number   Width:   8   Dec: 2
Variable: g3_lower_fence     Type: Number   Width:   8   Dec: 2
Variable: g3_upper_fence     Type: Number   Width:   8   Dec: 2
Variable: g3_boxplot_outlier   Type: Number   Width:   8   Dec: 2
Variable: g3_valid_n         Type: Number   Width:   7   Dec: 0
Variable: g3_outlier_count   Type: Number   Width:   8   Dec: 2
Variable: g3_outlier_percent   Type: Number   Width:   8   Dec: 2



* ------------------------------------------------------------.
* 11. Publishable SPSS GPL box-plot charts.
* ------------------------------------------------------------.

STRING g3_label (A30).
STRING absences_label (A30).
COMPUTE g3_label = "G3 final grade".
COMPUTE absences_label = "absences".
EXECUTE.

GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=g3_label G3
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: g3_label=col(source(s), name("g3_label"), unit.category())
 DATA: G3=col(source(s), name("G3"))
 GUIDE: text.title(label("Box Plot Interpretation: G3 Final Grade"))
 GUIDE: axis(dim(1), label(""))
 GUIDE: axis(dim(2), label("Final grade"))
 ELEMENT: schema(position(bin.quantile.letter(g3_label*G3)))
END GPL.
```

GGraph




```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=school_group G3
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: school_group=col(source(s), name("school_group"), unit.category())
 DATA: G3=col(source(s), name("G3"))
 GUIDE: text.title(label("G3 Box Plot by School"))
 GUIDE: axis(dim(1), label("School"))
 GUIDE: axis(dim(2), label("G3 final grade"))
 SCALE: cat(dim(1), values("1", "2"), labels("GP", "MS"))
 ELEMENT: schema(position(bin.quantile.letter(school_group*G3)))
END GPL.
```

GGraph

Warnings

GPL error: labels("GP","MS")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: school_group=col(source(s), name("school_group"),
unit.category())
DATA: G3=col(source(s), name("G3"))
GUIDE: text.title(label("G3 Box Plot by School"))
GUIDE: axis(dim(1), label("School"))
GUIDE: axis(dim(2), label("G3 final grade"))
SCALE: cat(dim(1), values("1", "2"), labels("GP", "MS"))
ELEMENT: schema(position(bin.quantile.letter
(school_group*G3)))

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=sex_group G3
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: sex_group=col(source(s), name("sex_group"), unit.category())
 DATA: G3=col(source(s), name("G3"))
 GUIDE: text.title(label("G3 Box Plot by Sex"))
 GUIDE: axis(dim(1), label("Sex"))
 GUIDE: axis(dim(2), label("G3 final grade"))
 SCALE: cat(dim(1), values("1", "2"), labels("Female", "Male"))
 ELEMENT: schema(position(bin.quantile.letter(sex_group*G3)))
END GPL.
```

GGraph

Warnings

GPL error: labels("Female","Male")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: sex_group=col(source(s), name("sex_group"), unit.
category())
DATA: G3=col(source(s), name("G3"))
GUIDE: text.title(label("G3 Box Plot by Sex"))
GUIDE: axis(dim(1), label("Sex"))
GUIDE: axis(dim(2), label("G3 final grade"))
SCALE: cat(dim(1), values("1", "2"), labels("Female", "Male"))
ELEMENT: schema(position(bin.quantile.letter
(sex_group*G3)))

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=studytime G3
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: studytime=col(source(s), name("studytime"), unit.category())
 DATA: G3=col(source(s), name("G3"))
 GUIDE: text.title(label("G3 Box Plot by Study Time"))
 GUIDE: axis(dim(1), label("Study time"))
 GUIDE: axis(dim(2), label("G3 final grade"))
 SCALE: cat(dim(1), values("1", "2", "3", "4"), labels("Less than 2 hours", "2
 to 5 hours", "5 to 10 hours", "More than 10 hours"))
 ELEMENT: schema(position(bin.quantile.letter(studytime*G3)))
END GPL.
```

GGraph

Warnings

GPL error: labels("Less than 2 hours","2 to 5 hours","5 to 10 hours","More than 10 hours")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: studytime=col(source(s), name("studytime"), unit.category())
DATA: G3=col(source(s), name("G3"))
GUIDE: text.title(label("G3 Box Plot by Study Time"))
GUIDE: axis(dim(1), label("Study time"))
GUIDE: axis(dim(2), label("G3 final grade"))
SCALE: cat(dim(1), values("1", "2", "3", "4"), labels("Less than 2 hours", "2 to 5 hours", "5 to 10 hours", "More than 10 hours"))
ELEMENT: schema(position(bin.quantile.letter(studytime*G3)))

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=failures G3
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: failures=col(source(s), name("failures"), unit.category())
 DATA: G3=col(source(s), name("G3"))
 GUIDE: text.title(label("G3 Box Plot by Past Failures"))
 GUIDE: axis(dim(1), label("Past failures"))
 GUIDE: axis(dim(2), label("G3 final grade"))
 SCALE: cat(dim(1), values("0", "1", "2", "3"), labels("0 failures", "1 failur
e", "2 failures", "3 failures"))
 ELEMENT: schema(position(bin.quantile.letter(failures*G3)))
END GPL.
```

GGraph

Warnings

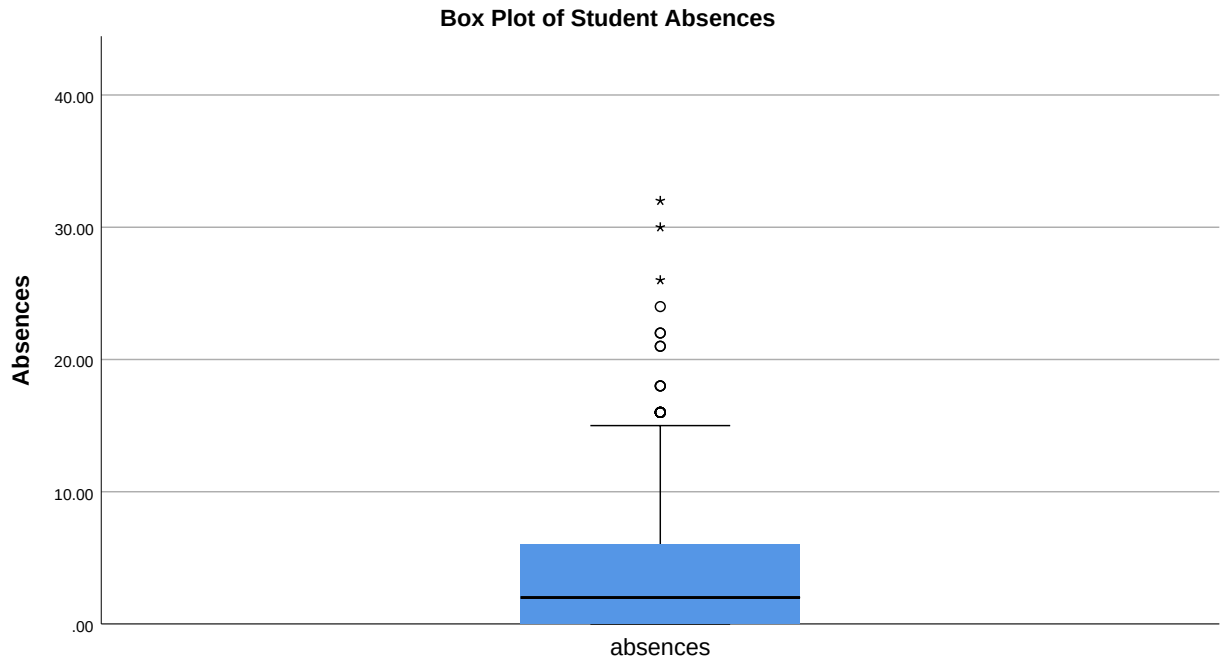
GPL error: labels("0 failures","1 failure","2 failures","3 failures")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: failures=col(source(s), name("failures"), unit.category())
DATA: G3=col(source(s), name("G3"))
GUIDE: text.title(label("G3 Box Plot by Past Failures"))
GUIDE: axis(dim(1), label("Past failures"))
GUIDE: axis(dim(2), label("G3 final grade"))
SCALE: cat(dim(1), values("0", "1", "2", "3"), labels("0 failures",
"1 failure", "2 failures", "3 failures"))
ELEMENT: schema(position(bin.quantile.letter(failures*G3)))

```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=absences_label absences
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: absences_label=col(source(s), name("absences_label"), unit.category())
 DATA: absences=col(source(s), name("absences"))
 GUIDE: text.title(label("Box Plot of Student Absences"))
 GUIDE: axis(dim(1), label(""))
 GUIDE: axis(dim(2), label("Absences"))
 ELEMENT: schema(position(bin.quantile.letter(absences_label*absences)))
END GPL.
```

GGraph




```
GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=school_group absences
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: school_group=col(source(s), name("school_group"), unit.category())
 DATA: absences=col(source(s), name("absences"))
 GUIDE: text.title(label("Absences Box Plot by School"))
 GUIDE: axis(dim(1), label("School"))
 GUIDE: axis(dim(2), label("Absences"))
 SCALE: cat(dim(1), values("1", "2"), labels("GP", "MS"))
 ELEMENT: schema(position(bin.quantile.letter(school_group*absences)))
END GPL.
```

GGraph

Warnings

GPL error: labels("GP","MS")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: school_group=col(source(s), name("school_group"), unit.category())
DATA: absences=col(source(s), name("absences"))
GUIDE: text.title(label("Absences Box Plot by School"))
GUIDE: axis(dim(1), label("School"))
GUIDE: axis(dim(2), label("Absences"))
SCALE: cat(dim(1), values("1", "2"), labels("GP", "MS"))
ELEMENT: schema(position(bin.quantile.letter (school_group*absences)))

```
* ------------------------------------------------------------.
* 12. G1, G2 and G3 comparison chart using long format.
* ------------------------------------------------------------.

DATASET COPY BoxPlotLong.
DATASET ACTIVATE BoxPlotLong.

VARSTOCASES
 /MAKE grade FROM G1 G2 G3
 /INDEX=grade_period
 /KEEP=case_id school_group sex_group studytime failures absences
 /NULL=DROP.
```

Variables to Cases

[BoxPlotLong]

Generated Variables

Name	Label
grade_period	<none>
grade	First period grade

Processing Statistics

Variables In	51
Variables Out	8

```
VARIABLE LABELS
 grade "Grade value"
 grade_period "Grade period".

VALUE LABELS grade_period
 1 "G1"
 2 "G2"
 3 "G3".

EXECUTE.

GGRAPH
 /GRAPHDATASET NAME="graphdataset" VARIABLES=grade_period grade
 /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
 SOURCE: s=userSource(id("graphdataset"))
 DATA: grade_period=col(source(s), name("grade_period"), unit.category())
 DATA: grade=col(source(s), name("grade"))
 GUIDE: text.title(label("Box Plot Comparison for G1, G2 and G3"))
 GUIDE: axis(dim(1), label("Grade period"))
 GUIDE: axis(dim(2), label("Grade"))
 SCALE: cat(dim(1), values("1", "2", "3"), labels("G1", "G2", "G3"))
 ELEMENT: schema(position(bin.quantile.letter(grade_period*grade)))
END GPL.
```

GGraph

Warnings

GPL error: labels("G1","G2","G3")

Undefined Function: labels

GPL:
SOURCE: s=userSource(id("graphdataset"))
DATA: grade_period=col(source(s), name("grade_period"), unit.category())
DATA: grade=col(source(s), name("grade"))
GUIDE: text.title(label("Box Plot Comparison for G1, G2 and G3"))
GUIDE: axis(dim(1), label("Grade period"))
GUIDE: axis(dim(2), label("Grade"))
SCALE: cat(dim(1), values("1", "2", "3"), labels("G1", "G2", "G3"))
ELEMENT: schema(position(bin.quantile.letter(grade_period*grade)))

```
DATASET ACTIVATE BoxPlotInterpretationClean.
DATASET CLOSE BoxPlotLong.

* ------------------------------------------------------------.
* 13. Save SPSS working file and corrected output PDF.
* ------------------------------------------------------------.

SAVE OUTFILE="D:\DATA ANALYSIS\A Basic Descriptive Statistics Guides\Box Plot
Interpretation\SPSS\box_plot_interpretation_spss_working_file_CORRECTED.sav"
 /COMPRESSED.

OUTPUT EXPORT
 /CONTENTS EXPORT=VISIBLE
```