PPGCOMP - FURG | 23148P - Data Visualization and Exploratory Data Analysis | 02/2024

This notebook contains the solution for Task 07 of the course 23148P - Data Visualization and Exploratory Data Analysis - 02/2024 of the Graduate Program in Computing at FURG (PPGCOMP-FURG).

Professor: Dr. Adriano Velasque Werhli.

Student: Vitor Avelaneda.

- **Contact:** avelaneda.vitor@gmail.com

The repository with the notebooks can be accessed here!

Task:

Readings

For this class you should read topics 12.1, 12.4 from here, and the whole chapter here.

Scatter plot

- **a.** See the scatter plots here. Create a dummy data set with 100 data points, $\{x \in N, 1 \leq x \leq 100\}$, where $y = \log(x) + N(\mu, \sigma^2)$ with $\mu - 0$ and $\sigma = 0.2$. In R the normal distribution $N(\mu, \sigma^2)$ is obtained with function rnorm.
  - **1.** Create a scatter plot with this data.
  - **2.** Add a linear trend using geom_smooth.
  - **3.** Add the confidence interval.
  - **4.** Add a LOESS trend with confidence interval.
  - **5.** For LOESS try to use the parameter `span=0.2`.
- **b.** Find a data set, or create one, for plotting a line graph similar to one of the examples here.

- **c.** Create a line plot where the x-axis represents time.

Solutions:

Verify the installation of necessary packages.

In [1]:
```
if (!requireNamespace("hrbrthemes", quietly = TRUE)) install.packages("hrbrthemes")
if (!requireNamespace("ggplot2", quietly = TRUE)) install.packages("ggplot2")
```

Load necessary packages.

In [2]:
```
library(hrbrthemes)
library(ggplot2)
```

> **a.** See the scatter plots here. Create a dummy data set with 100 data points, $\{x \in N, 1 \leq x \leq 100\}$, where $y = \log(x) + N(\mu, \sigma^2)$ with $\mu - 0$ and $\sigma = 0.2$. In R the normal distribution $N(\mu, \sigma^2)$ is obtained with function rnorm.

Solution...

In [3]:
```
df <- data.frame(
  x = 1:100,
  y = log(1:100) + rnorm(100, mean = 0, sd = 0.2)
)

View(df)
```

A data.frame: 100 × 2

x <int>	y <dbl>
1	-0.05791127
2	0.78343306
3	1.23689178
4	1.86636465
5	1.60704958
6	1.65642152
7	1.74583323
8	2.31825113
9	2.21028666
10	2.01022651
11	2.24745977
12	2.78356086
13	2.49021942
14	2.93561969
15	2.58127024
16	3.03106265
17	2.86069399
18	2.69641127
19	3.20003878
20	3.09602546
21	3.45521483

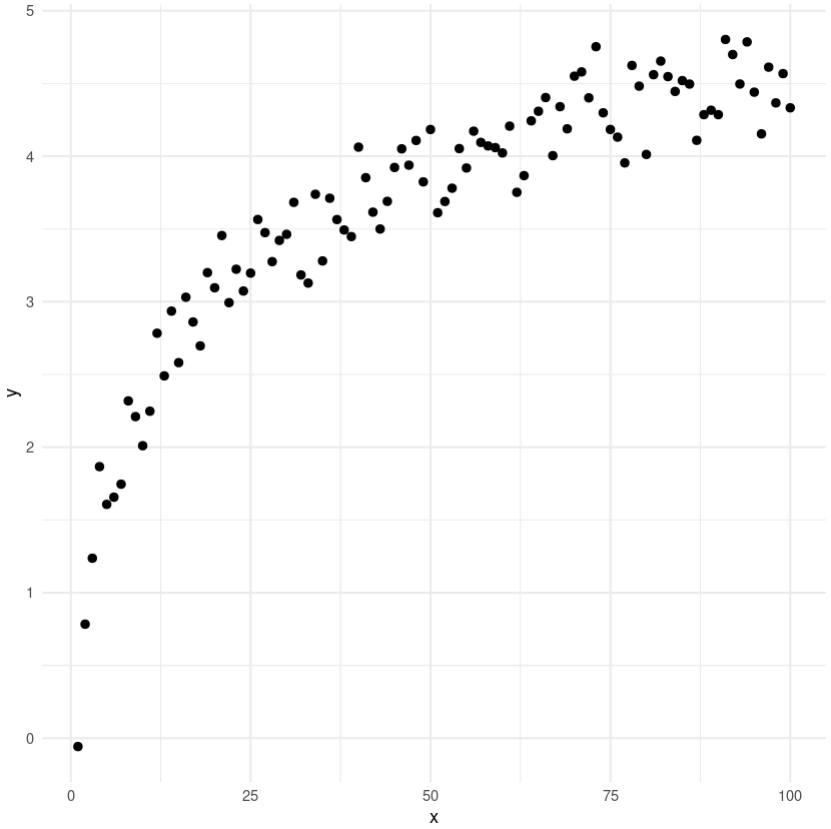
x	y
<int>	<dbl>
22	2.99372057
23	3.22343163
24	3.07372931
25	3.19675334
26	3.56515517
27	3.47495091
28	3.27558505
29	3.42161980
30	3.46342771
⋮	⋮
71	4.580032
72	4.401600
73	4.753325
74	4.298739
75	4.184048
76	4.131491
77	3.954505
78	4.624384
79	4.482060
80	4.012497
81	4.560848
82	4.654227

x	y
<int>	<dbl>
83	4.547682
84	4.445639
85	4.519993
86	4.496227
87	4.109558
88	4.286058
89	4.316442
90	4.286119
91	4.802761
92	4.699339
93	4.496732
94	4.786027
95	4.440669
96	4.154008
97	4.612514
98	4.366859
99	4.568201
100	4.333116

> **a. 1.** Create a scatter plot with this data.

Solution...

In [4]:
```
ggplot(df, aes(x = x, y = y)) +
  geom_point(color = "black", size = 2) +
  theme_minimal()
```

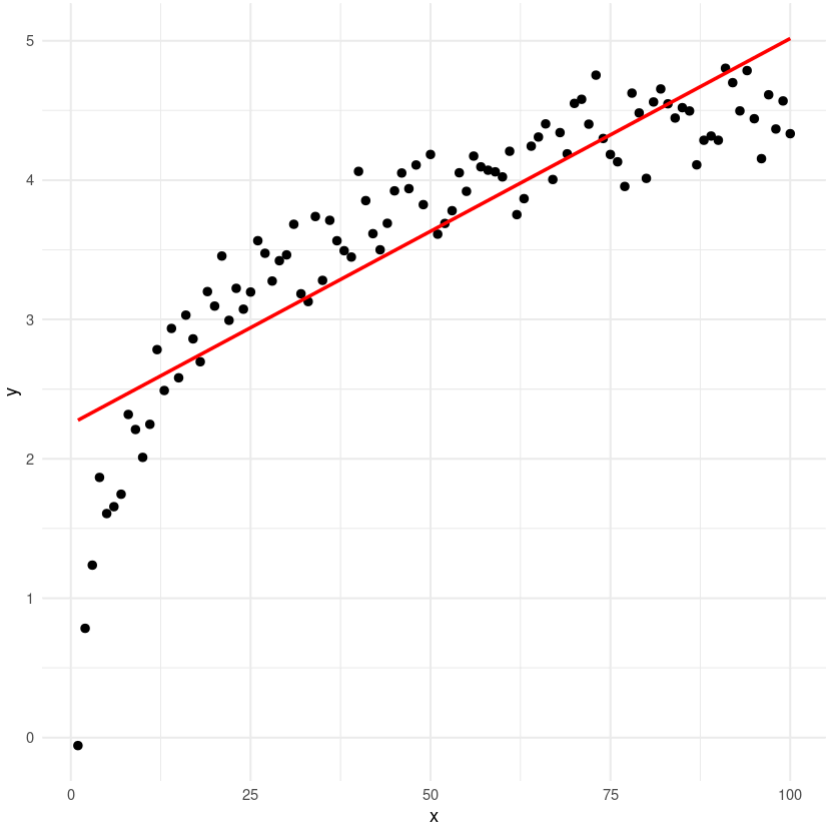


> **a. 2.** Add a linear trend using geom_smooth.

Solution...

In [5]:
```
ggplot(df, aes(x = x, y = y)) +
  geom_point(color = "black", size = 2) +
  geom_smooth(method = "lm", se = FALSE, color = "red", linetype = "solid") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

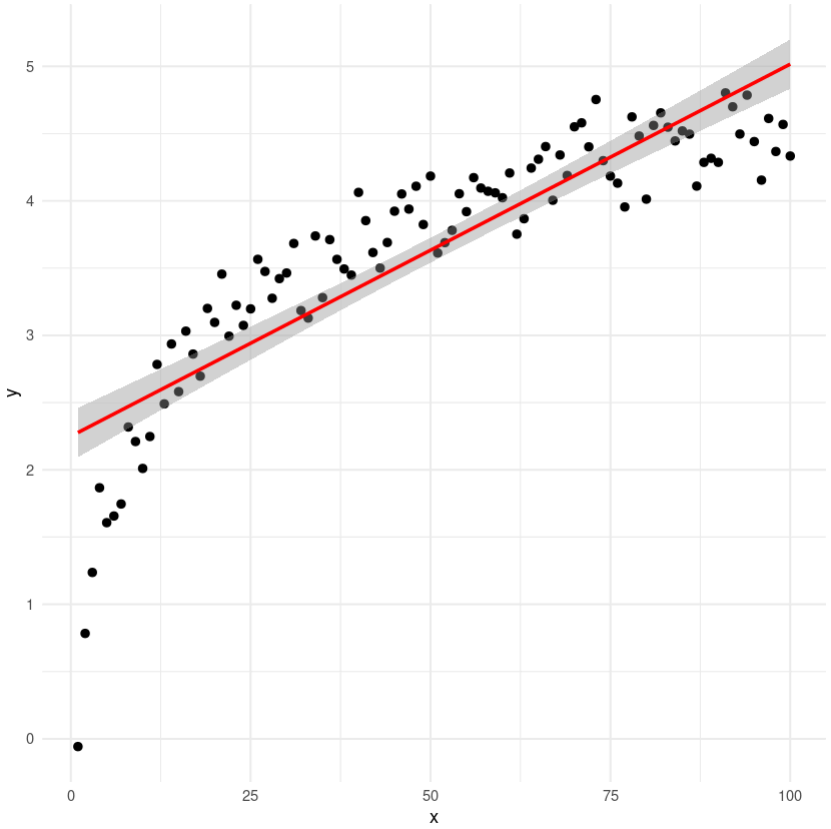


> **a. 3.** Add the confidence interval.

Solution...

In [6]:
```
ggplot(df, aes(x = x, y = y)) +
  geom_point(color = "black", size = 2) +
  geom_smooth(method = "lm", se = TRUE, color = "red", linetype = "solid") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

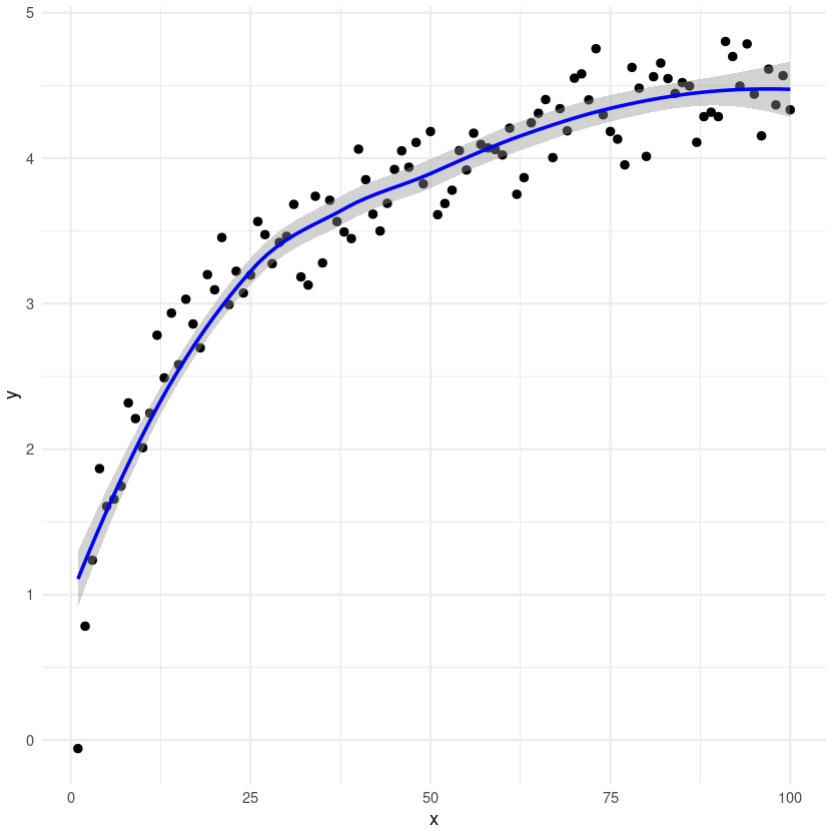


> **a. 4.** Add a LOESS trend with confidence interval.

Solution...

In [7]:
```
ggplot(df, aes(x = x, y = y)) +
  geom_point(color = "black", size = 2) +
  geom_smooth(method = "loess", se = TRUE, color = "blue", linetype = "solid") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

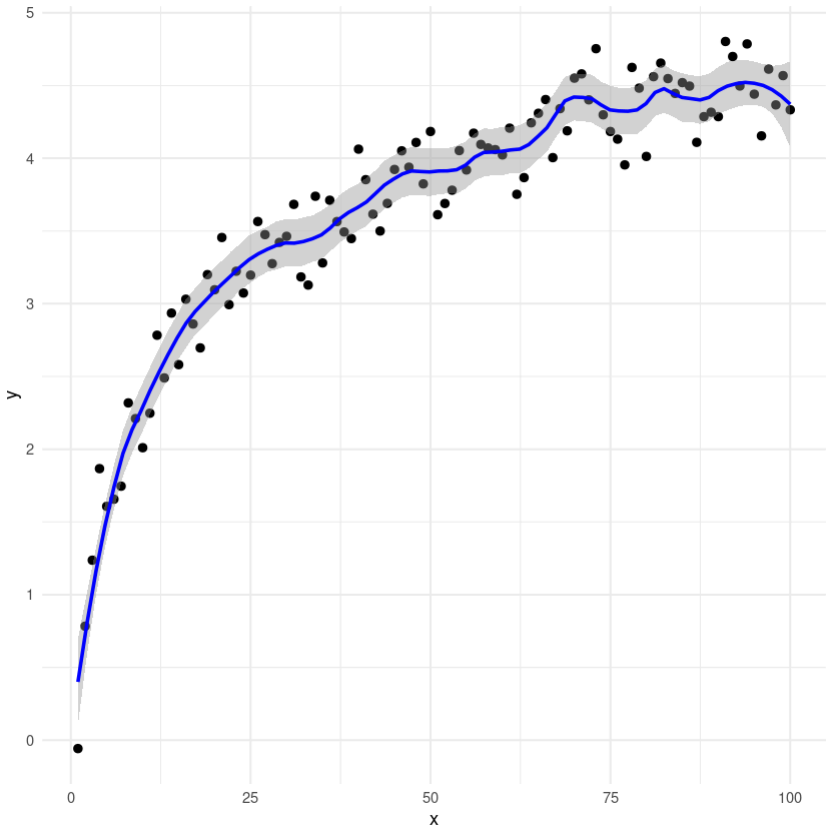


> **a. 5.** For LOESS try to use the parameter `span=0.2`.

Solution...

In [8]:
```
ggplot(df, aes(x = x, y = y)) +
  geom_point(color = "black", size = 2) +
  geom_smooth(method = "loess", span = 0.2, se = TRUE, color = "blue", linetype = "solid") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



> **b.** Find a data set, or create one, for plotting a line graph similar to one of the examples here.

Create data set...

In [9]:
```
set.seed(123)
df_ts <- data.frame(
  Time = seq(as.Date("2010-01-01"), by = "month", length.out = 100),
  Value = rnorm(100, mean = 50, sd = 5) + seq(-10, 10, length.out = 100)
)
View(df_ts)
```

A data.frame: 100 × 2

Time	Value
<date>	<dbl>
2010-01-01	37.19762
2010-02-01	39.05113
2010-03-01	48.19758
2010-04-01	40.95860
2010-05-01	41.45452
2010-06-01	49.58543
2010-07-01	43.51670
2010-08-01	35.08884
2010-09-01	38.18190
2010-10-01	39.58987
2010-11-01	48.14061
2010-12-01	44.02129
2011-01-01	44.42810
2011-02-01	43.17968
2011-03-01	40.04908
2011-04-01	51.96487
2011-05-01	45.72158
2011-06-01	33.60126
2011-07-01	47.14314
2011-08-01	41.47443
2011-09-01	38.70129
2011-10-01	43.15255

Time	Value
<date>	<dbl>
2011-11-01	39.31442
2011-12-01	41.00201
2012-01-01	41.72329
2012-02-01	36.61704
2012-03-01	49.44146
2012-04-01	46.22141
2012-05-01	39.96588
2012-06-01	52.12766
⋮	⋮
2015-11-01	51.68626
2015-12-01	42.79759
2016-01-01	59.57415
2016-02-01	51.20147
2016-03-01	51.50945
2016-04-01	60.27937
2016-05-01	53.92967
2016-06-01	49.45197
2016-07-01	56.66409
2016-08-01	55.26514
2016-09-01	56.19044
2016-10-01	58.29004
2016-11-01	54.71236

Time	Value
<date>	<dbl>
2016-12-01	59.98956
2017-01-01	55.86726
2017-02-01	58.83063
2017-03-01	62.85793
2017-04-01	59.75167
2017-05-01	56.14812
2017-06-01	63.72384
2017-07-01	63.14934
2017-08-01	61.12582
2017-09-01	59.77952
2017-10-01	55.64835
2017-11-01	65.79316
2017-12-01	56.19062
2018-01-01	70.33060
2018-02-01	67.25901
2018-03-01	58.61948
2018-04-01	54.86790

> **c.** Create a line plot where the x-axis represents time.

Solution...

In [10]:
```
ggplot(df_ts, aes(x = Time, y = Value)) +
  geom_line(color = "#69b3a2", size = 1) +
  labs(
```

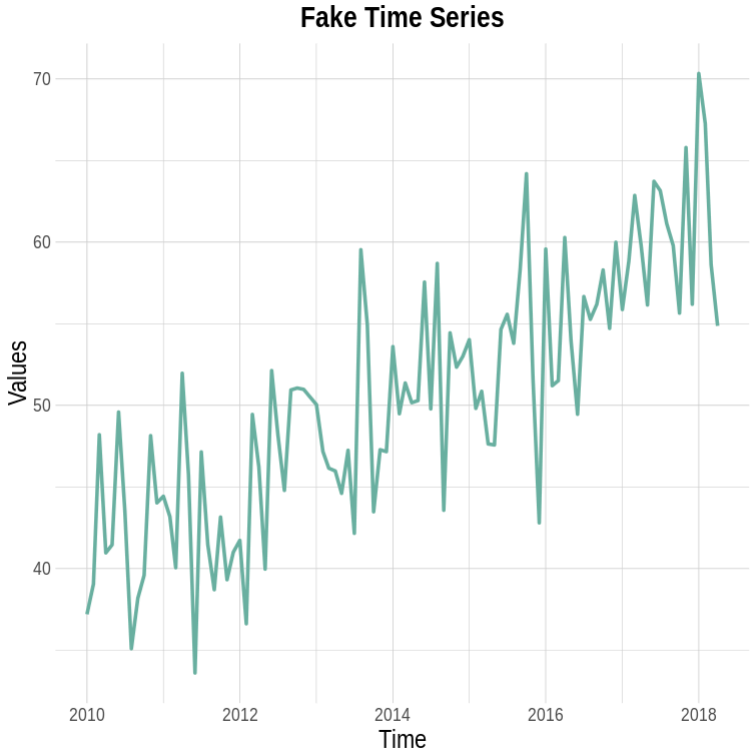
```
    title = "Fake Time Series",
    x = "Time",
    y = "Values"
  ) +
  theme_ipsum() +
  theme(
    axis.title.x = element_text(size = 16, hjust = 0.5),
    axis.title.y = element_text(size = 16, hjust = 0.5),
    plot.title = element_text(size = 18, hjust = 0.5)
  )
```

```
Warning message:
“Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
```

```
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
```

```
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)):
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
```

```
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
Warning message in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
“font family 'Arial Narrow' not found in PostScript font database”
```


Fake Time Series
70
60
50
40
Values
2010
2012
2014
2016
2018
Time