

M0130 – 3. PRAKTIKUM : M0130pr03 (*Úvodní analýza časových řad*)

A. Roční data:

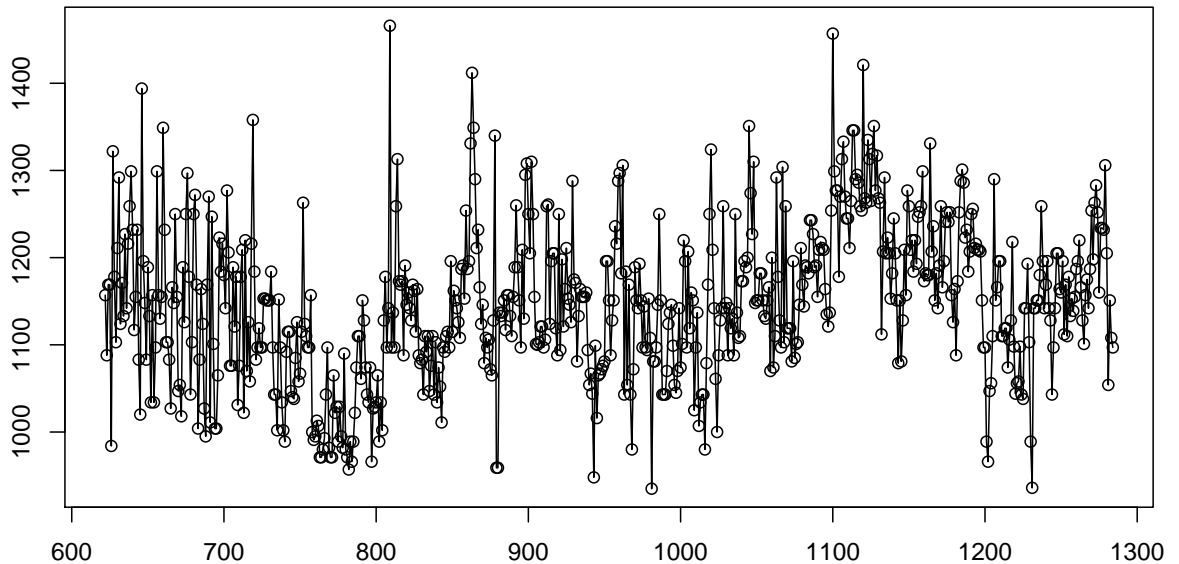
PŘÍKLAD 1

Máme k dispozici historická roční data, která se týkají minimálních (dolních) hladin řeky Nil, a to v letech 622–1284.

```
> Nile.low.levels <- c(1157, 1088, 1169, 1169, 984, 1322, 1178, 1103, 1211, 1292, 1124,
  1171, 1133, 1227, 1142, 1216, 1259, 1299, 1232, 1117, 1155, 1232, 1083, 1020, 1394,
  1196, 1148, 1083, 1189, 1133, 1034, 1157, 1034, 1097, 1299, 1157, 1130, 1155, 1349,
  1232, 1103, 1103, 1083, 1027, 1166, 1148, 1250, 1155, 1047, 1054, 1018, 1189, 1126,
  1250, 1297, 1178, 1043, 1103, 1250, 1272, 1169, 1004, 1083, 1164, 1124, 1027, 995,
  1169, 1270, 1011, 1247, 1101, 1004, 1004, 1065, 1223, 1184, 1216, 1180, 1142, 1277,
  1206, 1076, 1076, 1189, 1121, 1178, 1031, 1076, 1178, 1209, 1022, 1220, 1070, 1126,
  1058, 1216, 1358, 1184, 1083, 1097, 1119, 1097, 1097, 1153, 1153, 1151, 1151, 1151,
  1184, 1097, 1043, 1043, 1002, 1152, 1097, 1034, 1002, 989, 1092, 1115, 1115, 1047,
  1040, 1038, 1085, 1126, 1058, 1067, 1115, 1263, 1124, 1110, 1097, 1097, 1157, 1000,
  991, 995, 1013, 1007, 971, 971, 980, 993, 1043, 1097, 982, 971, 971, 1065, 1022,
  1029, 989, 1029, 995, 982, 1090, 980, 971, 957, 989, 966, 989, 1022, 1074, 1110,
  1110, 1061, 1151, 1128, 1074, 1043, 1034, 1074, 966, 1027, 1029, 1034, 1065, 989,
  1034, 1002, 1128, 1178, 1097, 1142, 1466, 1097, 1137, 1097, 1259, 1313, 1173, 1169,
  1173, 1088, 1191, 1146, 1160, 1142, 1128, 1169, 1162, 1115, 1164, 1088, 1079, 1083,
  1043, 1110, 1092, 1110, 1047, 1076, 1110, 1043, 1103, 1034, 1074, 1052, 1011, 1097,
  1092, 1110, 1115, 1097, 1196, 1115, 1162, 1151, 1142, 1126, 1108, 1187, 1191, 1153,
  1254, 1187, 1196, 1331, 1412, 1349, 1290, 1211, 1232, 1166, 1124, 1146, 1079, 1108,
  1097, 1106, 1072, 1065, 1128, 1340, 959, 959, 1137, 1133, 1137, 1151, 1117, 1157,
  1157, 1133, 1110, 1155, 1189, 1260, 1189, 1151, 1097, 1209, 1130, 1295, 1308, 1250,
  1205, 1310, 1250, 1155, 1101, 1100, 1103, 1121, 1121, 1097, 1106, 1259, 1261, 1124,
  1196, 1205, 1205, 1119, 1088, 1250, 1094, 1198, 1121, 1164, 1211, 1153, 1146, 1126,
  1288, 1175, 1171, 1081, 1133, 1164, 1155, 1155, 1155, 1160, 1094, 1054, 1067, 1044,
  948, 1099, 1016, 1065, 1067, 1072, 1076, 1081, 1196, 1196, 1151, 1088, 1128, 1151,
  1236, 1216, 1288, 1297, 1182, 1306, 1043, 1184, 1054, 1169, 1043, 980, 1072, 1189,
  1151, 1142, 1193, 1151, 1097, 1144, 1097, 1094, 1153, 1108, 935, 1081, 1081, 1097,
  1146, 1250, 1151, 1043, 1043, 1043, 1070, 1124, 1137, 1146, 1099, 1054, 1045, 1070,
  1142, 1074, 1101, 1220, 1196, 1097, 1207, 1119, 1160, 1151, 1025, 1097, 1137, 1007,
  1034, 1043, 1043, 980, 1079, 1169, 1250, 1324, 1209, 1142, 1061, 1000, 1088, 1128,
  1142, 1259, 1142, 1148, 1088, 1142, 1119, 1130, 1088, 1250, 1137, 1108, 1110, 1173,
  1173, 1196, 1189, 1200, 1351, 1274, 1227, 1310, 1148, 1151, 1151, 1182, 1182, 1151,
  1133, 1130, 1151, 1166, 1070, 1200, 1074, 1110, 1292, 1178, 1128, 1097, 1304, 1103,
  1259, 1119, 1119, 1119, 1081, 1196, 1085, 1101, 1103, 1146, 1211, 1169, 1144, 1191,
  1189, 1182, 1243, 1243, 1227, 1189, 1191, 1155, 1209, 1218, 1211, 1209, 1164, 1135,
  1121, 1137, 1254, 1457, 1299, 1277, 1277, 1178, 1270, 1313, 1333, 1270, 1245, 1245,
  1211, 1265, 1346, 1346, 1290, 1295, 1286, 1259, 1254, 1421, 1268, 1263, 1335, 1313,
  1265, 1319, 1351, 1277, 1317, 1268, 1263, 1112, 1207, 1292, 1205, 1223, 1205, 1153,
  1182, 1245, 1205, 1151, 1079, 1151, 1081, 1128, 1209, 1157, 1277, 1259, 1209, 1220,
  1184, 1220, 1193, 1247, 1252, 1259, 1299, 1173, 1182, 1180, 1180, 1331, 1207, 1236,
  1151, 1182, 1142, 1191, 1259, 1166, 1196, 1241, 1252, 1241, 1252, 1157, 1126, 1164,
  1088, 1173, 1252, 1288, 1301, 1286, 1223, 1232, 1184, 1207, 1250, 1256, 1211, 1216,
  1209, 1209, 1207, 1151, 1097, 1097, 989, 966, 1047, 1056, 1110, 1290, 1151, 1166,
  1196, 1196, 1110, 1110, 1119, 1119, 1074, 1106, 1128, 1218, 1098, 1044, 1056, 1058,
  1098, 1043, 1038, 1142, 1142, 1193, 1103, 989, 936, 1142, 1142, 1151, 1151, 1180,
  1259, 1196, 1142, 1169, 1196, 1142, 1128, 1043, 1097, 1142, 1205, 1205, 1164, 1160,
  1196, 1112, 1169, 1110, 1178, 1133, 1153, 1139, 1155, 1187, 1196, 1220, 1166, 1128,
  1101, 1157, 1175, 1142, 1187, 1254, 1198, 1263, 1283, 1252, 1160, 1234, 1234, 1232,
  1306, 1205, 1054, 1151, 1108, 1097)
```

Vytvoříme časovou řadu, kterou vykreslíme.

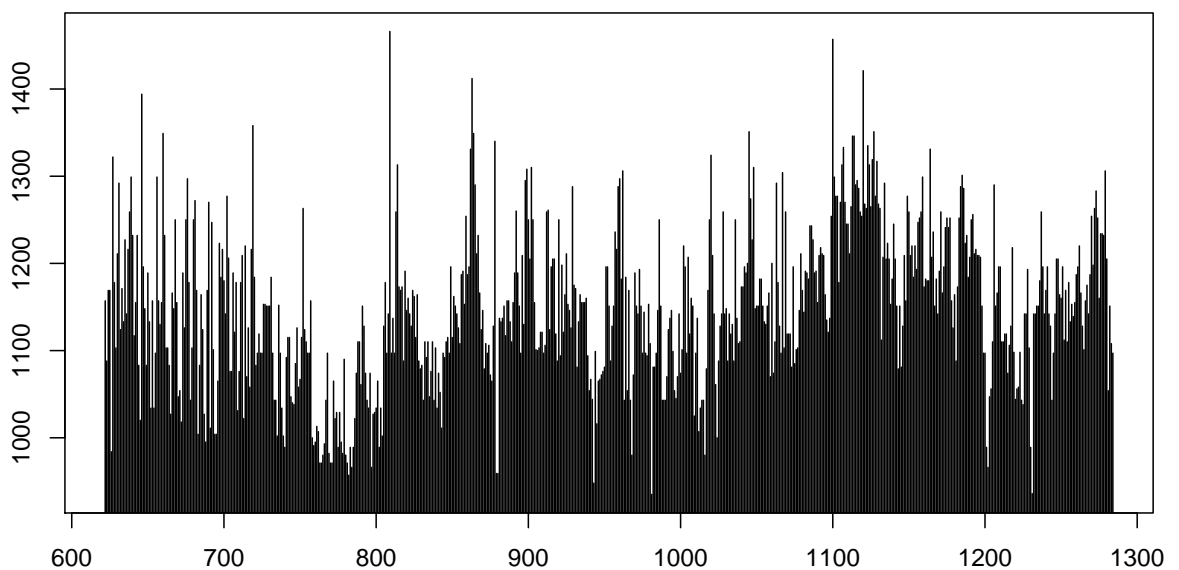
```
> NileTS <- ts(nile.low.levels, start = 622, frequency = 1)
> par(mar = c(2, 2, 0, 0) + 0.5)
> plot(NileTS, type = "o")
```



Obrázek 1: Minimální roční hladiny řeky Nil v letech 622–1284 (Roda Island, Cairo)

Vyzkoušíme, jak bude vypadat graf časové řady s volbou `type="h"`

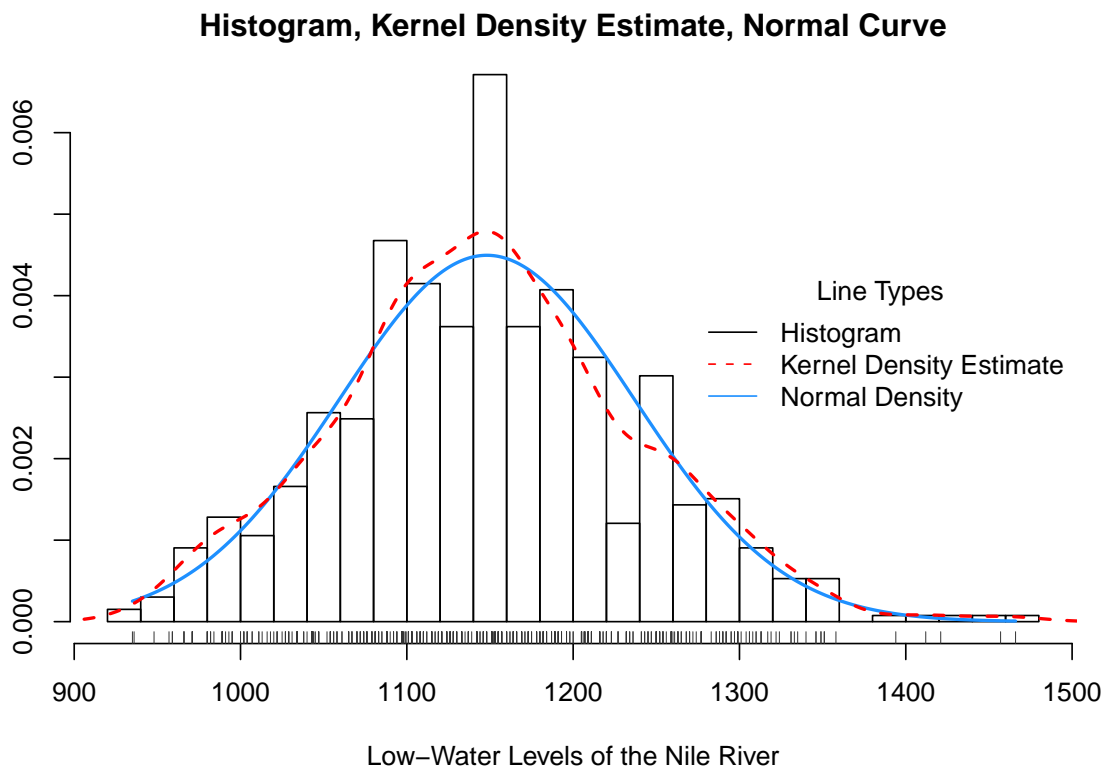
```
> par(mar = c(2, 2, 0, 0) + 0.5)
> plot(NileTS, type = "h")
```



Obrázek 2: Minimální roční hladiny řeky Nil v letech 622–1284 - s volbou `type="h"`

Nyní vytvoříme histogram, jádrový odhad hustoty vstupních dat a navíc ještě vykreslíme normální hustotu, kde střední hodnota a rozptyl jsou určeny výběrovými odhady.

```
> x <- NileTS
> par(mar = c(4, 2, 1, 0) + 0.75)
> h <- hist(x, probability = TRUE, breaks = "FD", xlab = "Low-Water Levels of the Nile River",
  main = "Histogram, Kernel Density Estimate, Normal Curve")
> xfit <- seq(min(x), max(x), length = 512)
> yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
> lines(xfit, yfit, col = "dodgerblue", lwd = 2)
> lines(density(x, n = 512), lwd = 2, col = "red", lty = 2)
> rug(x, side = 1, ticksize = 0.02, col = "grey20")
> legend("right", legend = c("Histogram", "Kernel Density Estimate",
  "Normal Density"), col = c("black", "red", "dodgerblue"),
  lty = c(1, 2, 1), bty = "n", title = "Line Types")
```



Obrázek 3: Histogram, jádrový odhad hustoty, normální hustota pro roční minima

B. Měsíční data:

PŘÍKLAD 2

V tomto případě máme k dispozici měsíční data, která udávají počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002.

Nejprve načteme popis časové řady, který je v datovém souboru umístěn na prvních dvou rádcích. Můžeme to udělat několika způsoby.

První možností je použít příkaz `readLines`. Protože je příkaz v závorkách, ihned se zobrazí obsah proměnné `popis`.

```
> fileDat <- paste(data.library, "MontlyUnemplWomen.dat",
  sep = "")
> con <- file(fileDat)
> (popis <- readLines(con, n = 2))

[1] "Monthly unemployed young women between ages 16 and 19"
[2] "in the U.S. from January 1961 to August 2002"

> close(con)
```

Další možností je použít příkaz `scan`.

```
> (txt <- scan(fileDat, what = "", nlines = 2))

 [1] "Monthly"      "unemployed" "young"      "women"      "between"
 [6] "ages"         "16"         "and"        "19"         "in"
[11] "the"          "U.S."       "from"       "January"    "1961"
[16] "to"           "August"    "2002"
```

Vidíme, že každé slovo se načetlo zvlášť, chceme-li všechna slova spojit v jediný text, píšeme

```
> (Txt <- paste(txt, collapse = " "))

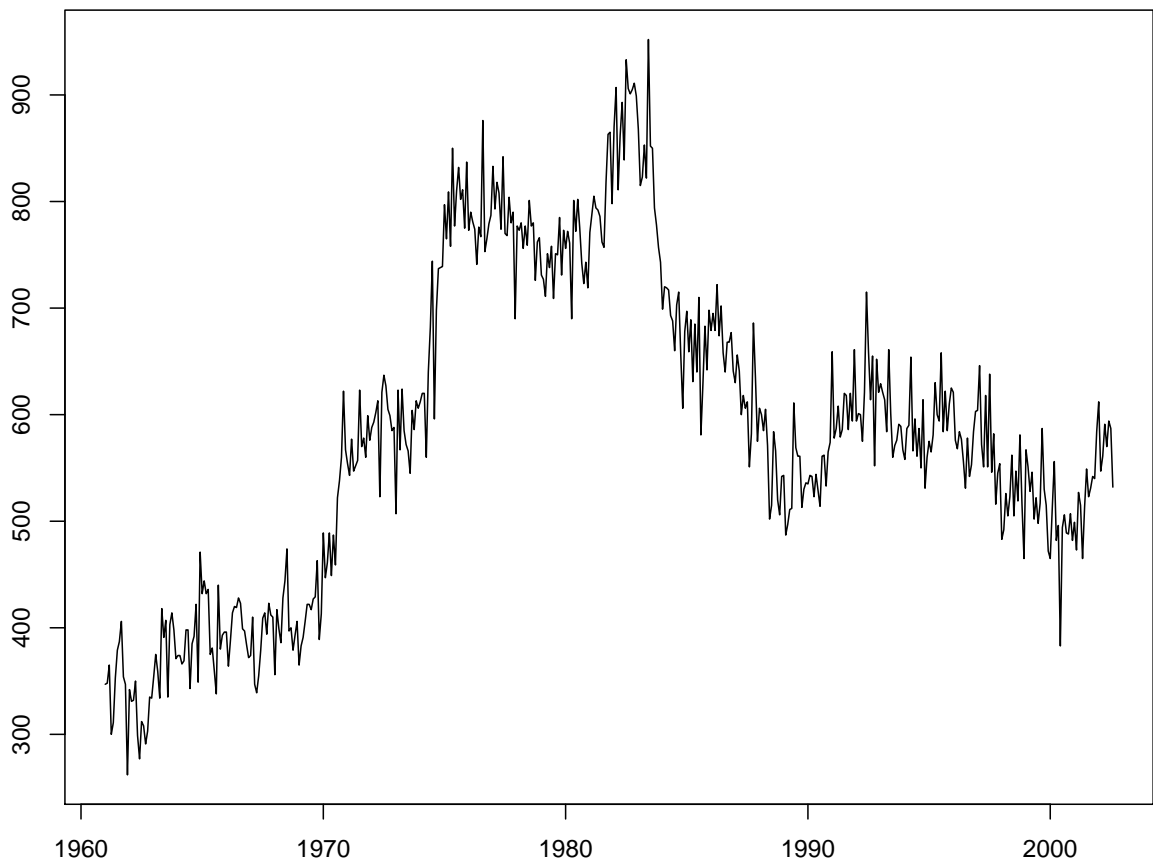
[1] "Monthly unemployed young women between ages 16 and 19 in the U.S. from January 1961 to August 2002"
```

Nyní načteme zbytek dat pomocí příkazu `scan` a volbou `skip3` zajistíme přeskočení prvních dvou řádků s textem a jednoho volného řádku.

```
> UnemplWomen <- scan(fileDat, skip = 3)
```

Z načtených dat vytvoříme časovou řadu a vykreslíme ji.

```
> UnemplWomenTS <- ts(UnemplWomen, start = 1961, frequency = 12)
> par(mar = c(2, 2, 0, 0) + 0.5)
> plot(UnemplWomenTS)
```



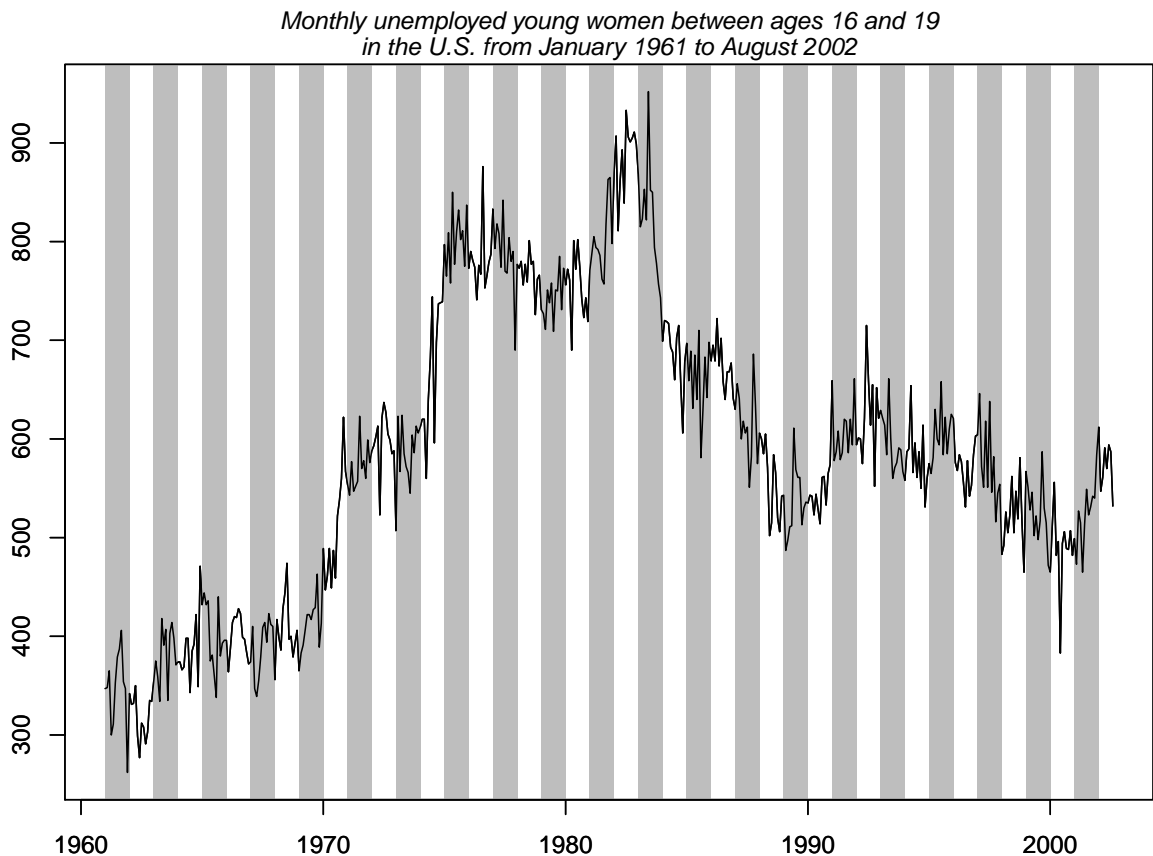
Obrázek 4: Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002

Abychom mohli používat graf, ve kterém budou zřetelně odděleny roční bloky dat, nejprve vytvoříme funkci

```
> plot.band <- function(x, lwd = 1.5, type = "o", pch = 20,
... ) {
  plot(x, ...)
  a <- time(x)
  i1 <- floor(min(a))
  i2 <- ceiling(max(a))
  y1 <- par("usr")[3]
  y2 <- par("usr")[4]
  if (par("ylog")) {
    y1 <- 10^y1
    y2 <- 10^y2
  }
  for (i in seq(from = i1, to = i2 - 1, by = 2)) {
    polygon(c(i, i + 1, i + 1, i), c(y1, y1, y2,
      y2), col = "grey", border = NA)
  }
  par(new = T)
  plot(x, lwd = lwd, type = type, pch = pch, ...)
}
```

Tuto funkci pak použijeme pro grafické oddělení jednotlivých let.

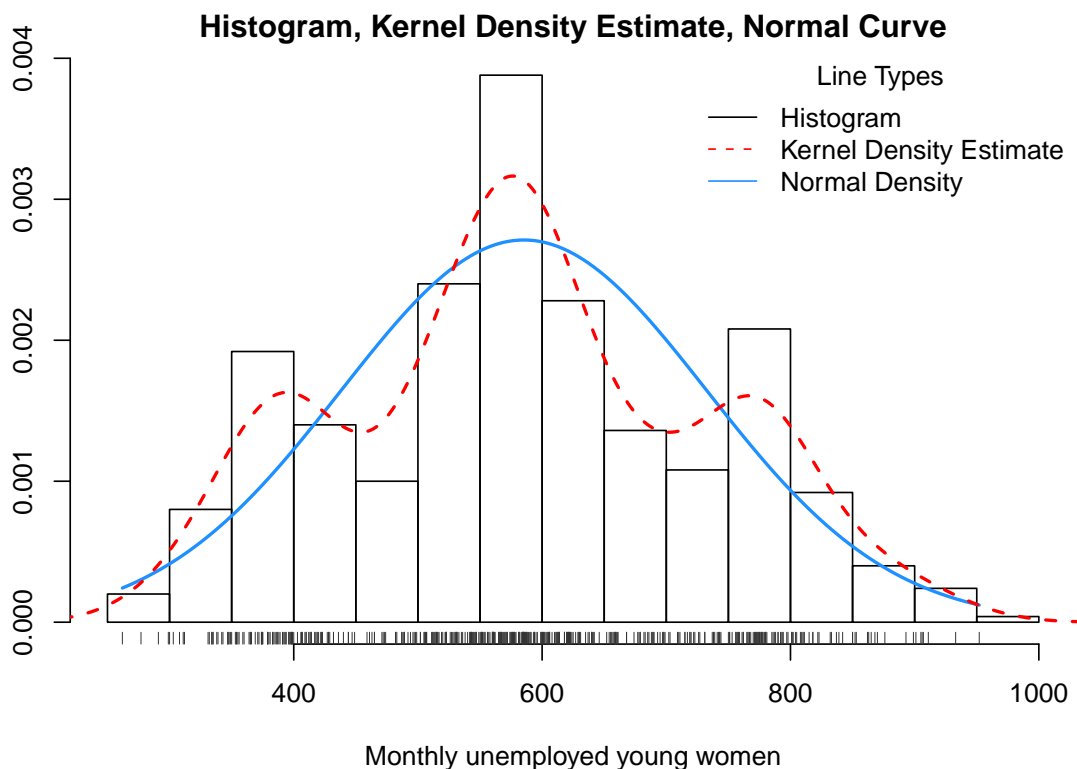
```
> par(mar = c(2, 2, 2, 0) + 0.5)
> plot.band(UnemplWomenTS, type = "l", lwd = 1)
> mtext(popis[1], line = 1, font = 3)
> mtext(popis[2], line = 0.125, font = 3)
```



Obrázek 5: Počet nezaměstnaných mladých žen ve věku od 16 do 19 let v USA od ledna 1961 do srpna 2002 – roční bloky

Nyní vytvoříme histogram, jádrový odhad hustoty vstupních dat a navíc ještě vykreslíme normální hustotu, kde střední hodnota a rozptyl jsou určeny výběrovými odhady.

```
> x <- UnemplWomenTS
> par(mar = c(4, 2, 1, 0) + 0.75)
> h <- hist(x, probability = TRUE, breaks = "FD", xlab = "Monthly unemployed young women",
  main = "Histogram, Kernel Density Estimate, Normal Curve")
> xfit <- seq(min(x), max(x), length = 512)
> yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
> lines(xfit, yfit, col = "dodgerblue", lwd = 2)
> lines(density(x, n = 512), lwd = 2, col = "red", lty = 2)
> rug(x, side = 1, ticksize = 0.02, col = "grey20")
> legend("topright", legend = c("Histogram", "Kernel Density Estimate",
  "Normal Density"), col = c("black", "red", "dodgerblue"),
  lty = c(1, 2, 1), bty = "n", title = "Line Types")
```



Obrázek 6: Histogram, jádrový odhad hustoty, normální hustota pro počty mladých nezaměstnaných žen

Jak je vidět z předchozího grafu, vstupní data jsou směsí tří rozdělení.

C. Data s desetinnou čárkou:

PŘÍKLAD 3

Datový soubor `ZiveNarozeni.dat` obsahuje v prvním řádku popisku dat, druhý řádek je volný, od třetího řádku jsou k dispozici data, ve kterých je místo desetinné tečky desetinná čárka.

Nejprve načteme popis datového souboru.

```
> fileDat <- paste(data.library, "ZiveNarozeni.dat", sep = "")
> con <- file(fileDat)
> (popis <- readLines(con, n = 1))
```

```
[1] "ČR: Živě narození na 1000 obyvatel v letech 1785 - 2009"
```

```
> close(con)
```

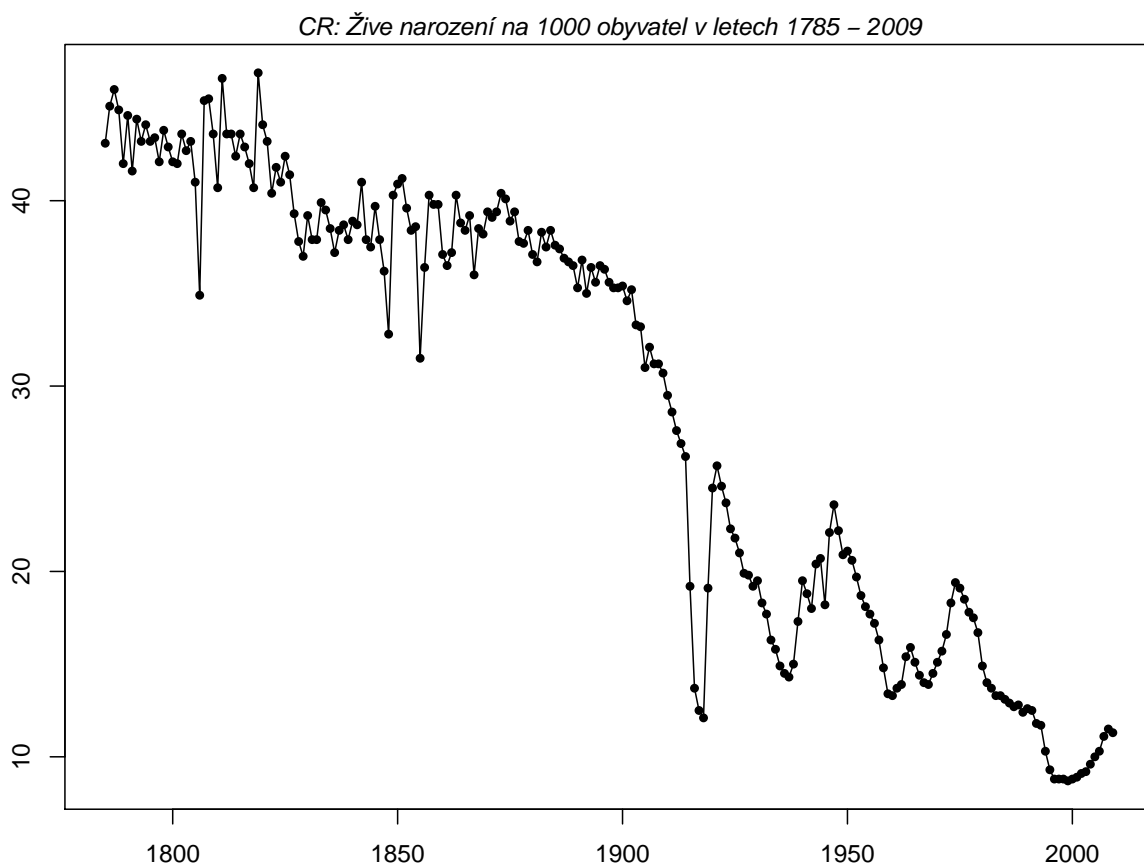
Nyní načteme samotná data pomocí příkazu `scan`. Volba `skip=2` zajistí přeskočení popisného a volného řádku a volba `dec=", "` umožní načíst číslo s desetinnou čárkou. Pokud bychom to neudělali, příkaz `scan` by vstupní data bral jako typu `character`.

```
> ZiveNarozeni <- scan(fileDat, skip = 2, dec = ",")
> str(ZiveNarozeni)
```

```
num [1:225] 43.1 45.1 46 44.9 42 44.6 41.6 44.4 43.2 44.1 ...
```

Z načtených dat vytvoříme časovou řadu a vykreslíme ji.

```
> ZiveNarozeniTS <- ts(ZiveNarozeni, start = 1785, frequency = 1)
> par(mar = c(2, 2, 1, 0) + 0.5)
> plot(ZiveNarozeniTS, type = "o", pch = 20)
> mtext(popis, line = 0.125, font = 3)
```



Obrázek 7: Počet živě narozených dětí na 1000 obyvatel v České republice v letech 1785 - 2009.

Nyní vytvoříme histogram, jádrový odhad hustoty vstupních dat a navíc ještě vykreslíme normální hustotu, kde střední hodnota a rozptyl jsou určeny výběrovými odhady.

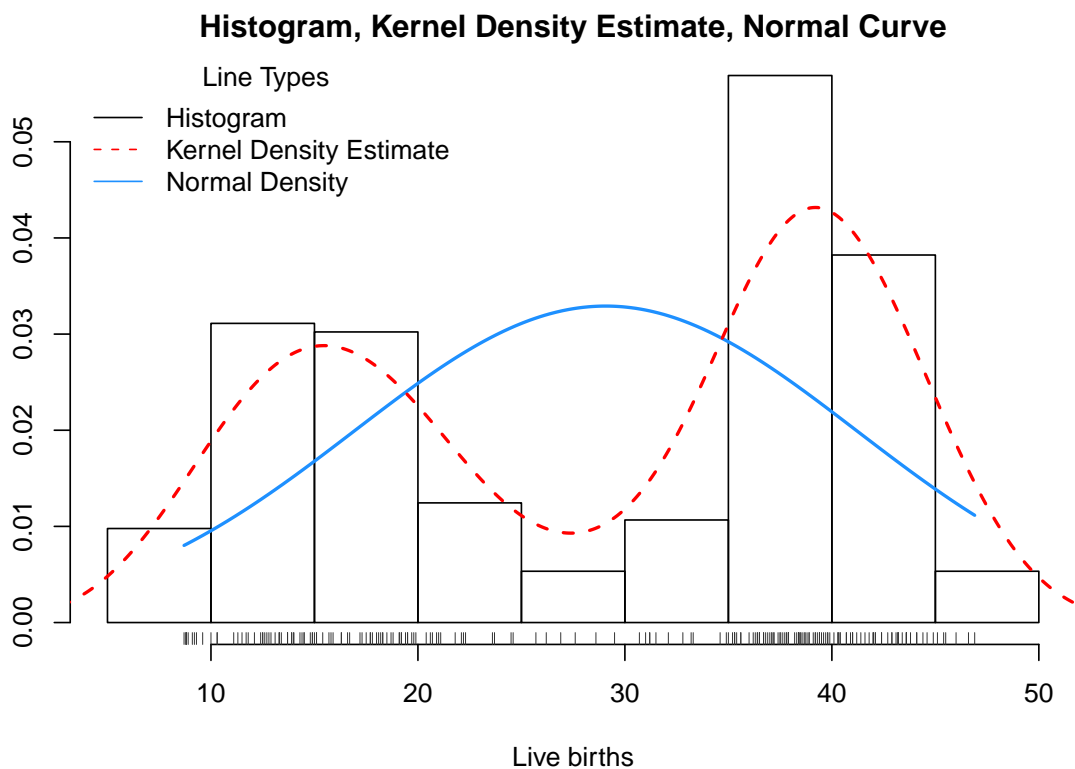
```
> x <- ZiveNarozeniTS
> par(mar = c(4, 2, 1, 0) + 0.75)
> h <- hist(x, probability = TRUE, breaks = "FD", xlab = "Live births",
  main = "Histogram, Kernel Density Estimate, Normal Curve")
> xfit <- seq(min(x), max(x), length = 512)
> yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
```



```

> lines(xfit, yfit, col = "dodgerblue", lwd = 2)
> lines(density(x, n = 512), lwd = 2, col = "red", lty = 2)
> rug(x, side = 1, ticksize = 0.02, col = "grey20")
> legend("topleft", legend = c("Histogram", "Kernel Density Estimate",
  "Normal Density"), col = c("black", "red", "dodgerblue"),
  lty = c(1, 2, 1), bty = "n", title = "Line Types")

```



Obrázek 8: Histogram, jádrový odhad hustoty, normální hustota pro počty živě narozených dětí.

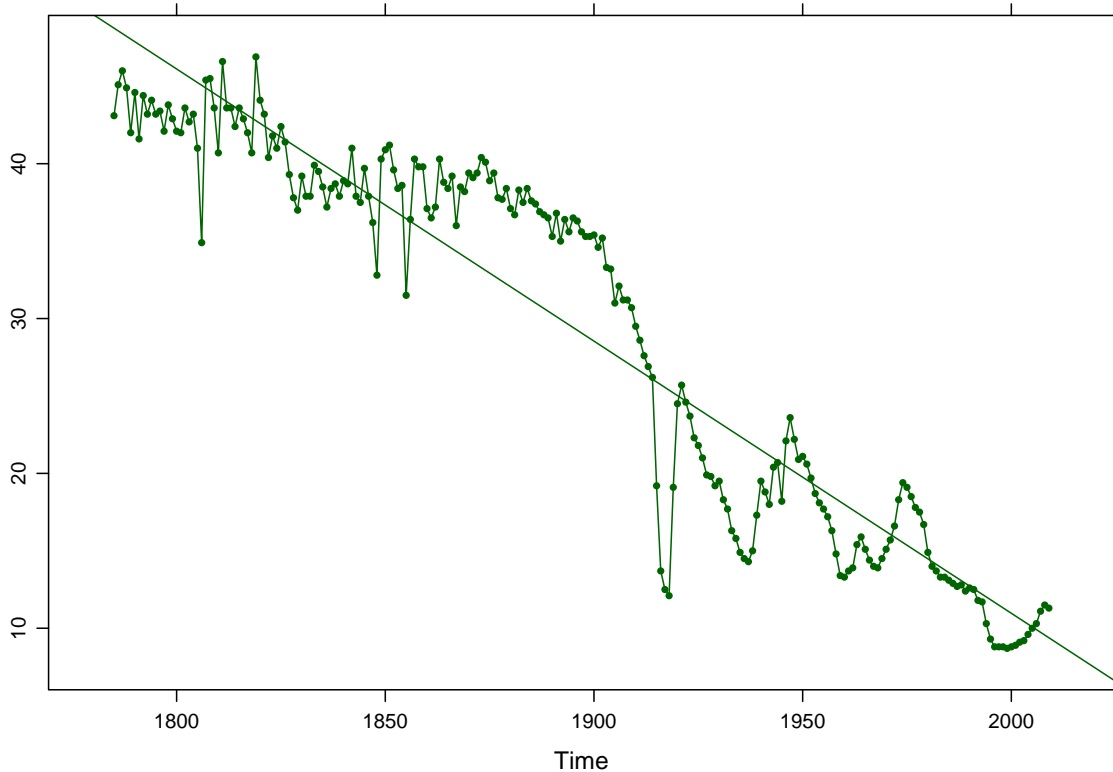
K vykreslení dat můžeme použít i grafy z knihovny `lattice`, kdy máme možnost rozšířit volbu `type=c("o", "r")` tak, aby se navíc daty proložila regresní přímka.

```

> library(lattice)
> print(xyplot(ZiveNarozeniTS, type = c("o", "r"), pch = 20,
  main = popis))

```

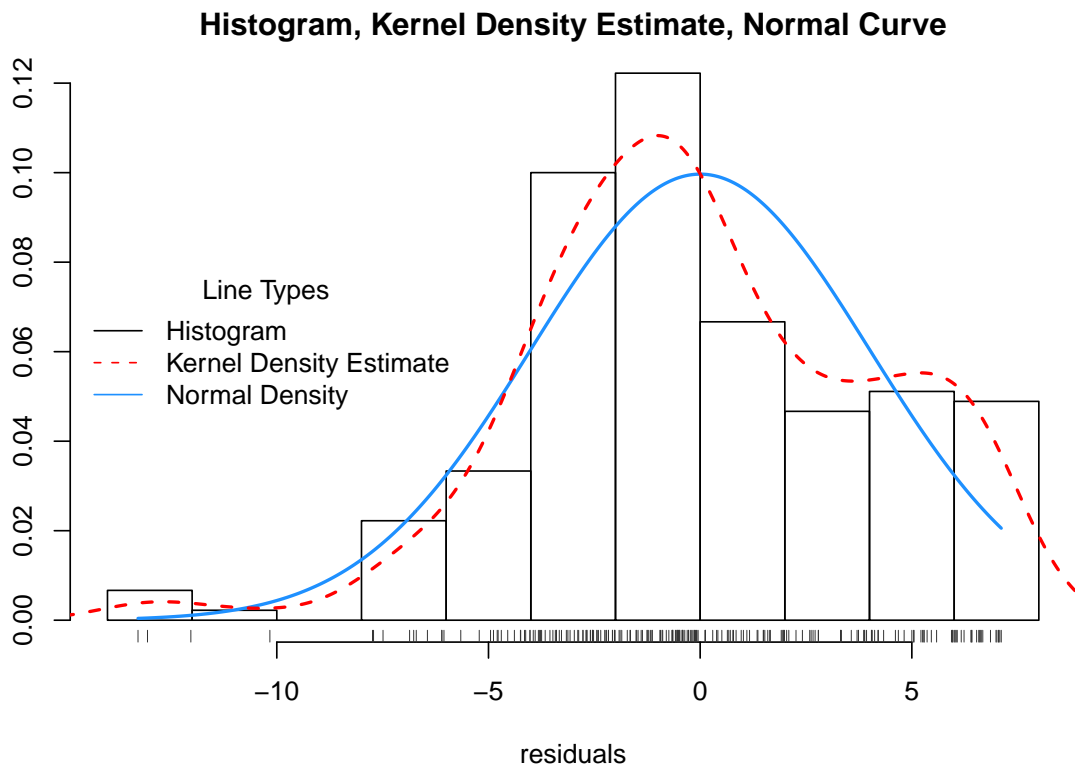
CR: Živé narození na 1000 obyvatel v letech 1785 – 2009



Obrázek 9: Počet živě narozených dětí na 1000 obyvatel v České republice v letech 1785 - 2009 spolu s regresní přímkou.

Na závěr ještě vyšetříme, jak dopadnou odhady rozdělení, pokud z časové řady odstraníme lineární trend, tj. vykreslíme histogram, jádrový odhad hustoty a normální hustotu pro rezidua, která získáme z jednoduchého regresního modelu s lineárním trendem.

```
> Time <- 1:length(ZiveNarozeni)
> x <- residuals(lm(ZiveNarozeni ~ Time))
> par(mar = c(4, 2, 1, 0) + 0.75)
> h <- hist(x, probability = TRUE, breaks = "FD", xlab = "residuals",
  main = "Histogram, Kernel Density Estimate, Normal Curve")
> xfit <- seq(min(x), max(x), length = 512)
> yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
> lines(xfit, yfit, col = "dodgerblue", lwd = 2)
> lines(density(x, n = 512), lwd = 2, col = "red", lty = 2)
> rug(x, side = 1, ticksize = 0.02, col = "grey20")
> legend("left", legend = c("Histogram", "Kernel Density Estimate",
  "Normal Density"), col = c("black", "red", "dodgerblue"),
  lty = c(1, 2, 1), bty = "n", title = "Line Types")
```



Obrázek 10: Histogram, jádrový odhad hustoty, normální hustota pro residua.

Srovnáme-li obrázky 8 a 10, vidíme, že pokud je u časových řad výrazný trend, nemá smysl odhadovat rozdělení dat, neboť v nich velkou váhu má deterministická složka. Tu je třeba nejprve odstranit a pak teprve zkoumat rozdělení zbývajících stochastických složek.

D. Úkol:

Pro časové řady, které jste si našli v minulém praktiku, proveďte úvodní analýzu.