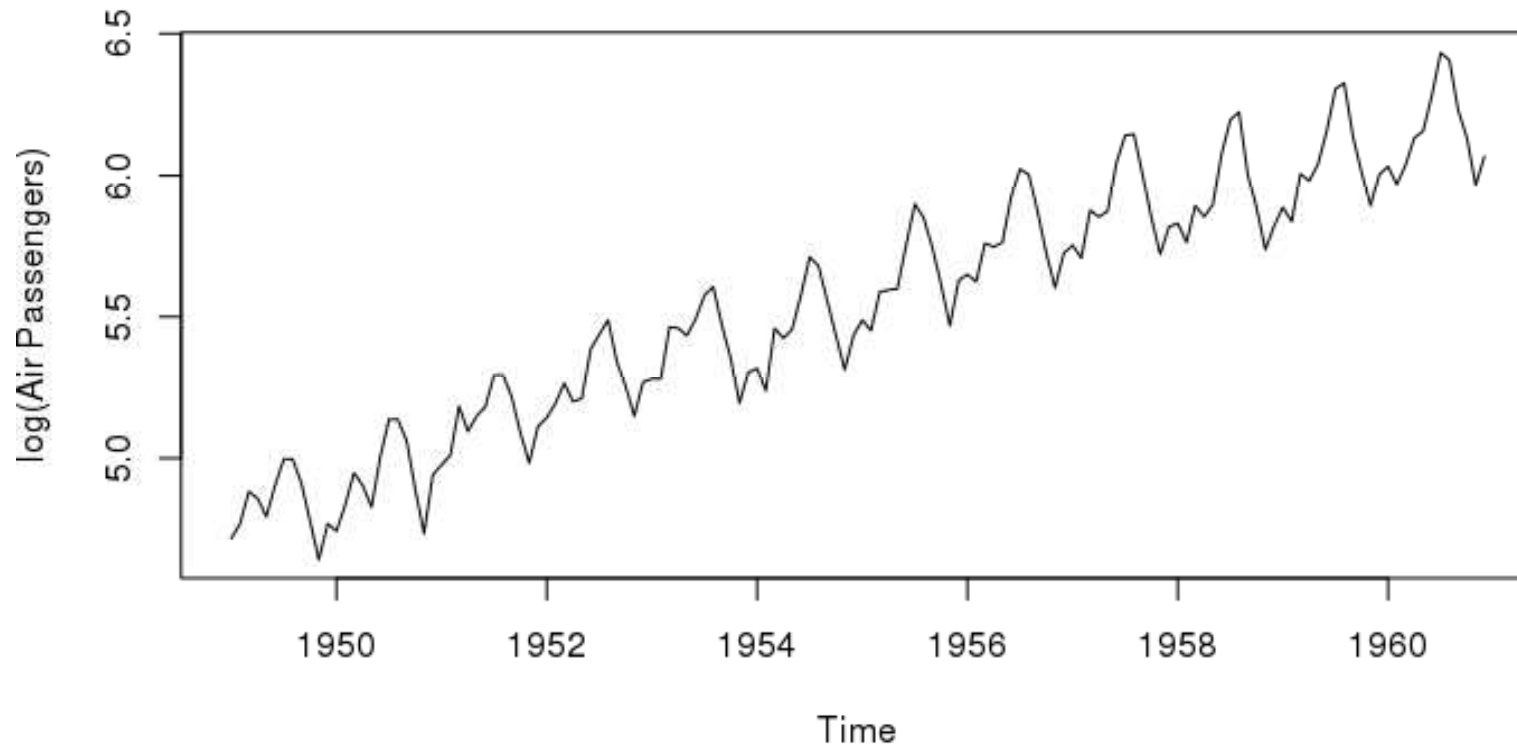


Spektrálna analýza

Beáta Stehlíková
Časové rady, FMFI UK

Motivácia

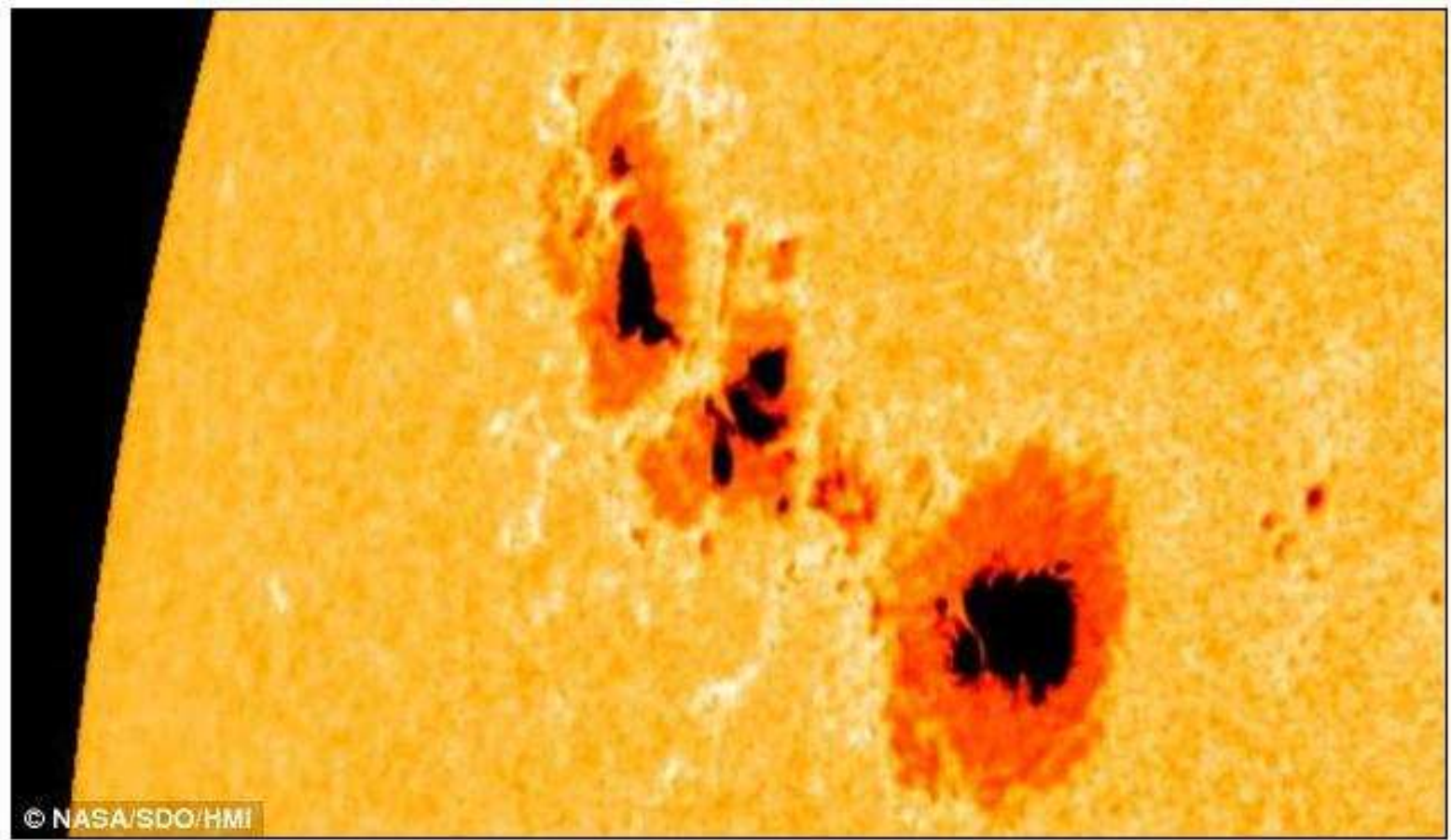
- Príklad o počte cestujúcich aerolinkami:



- Podobne napr. mesačné dáta nezamestnanosti, kvartálne dáta HDP, ... → *modely na cvičení*
- Nie vždy tam však musí byť takto jasná perióda

Motivácia

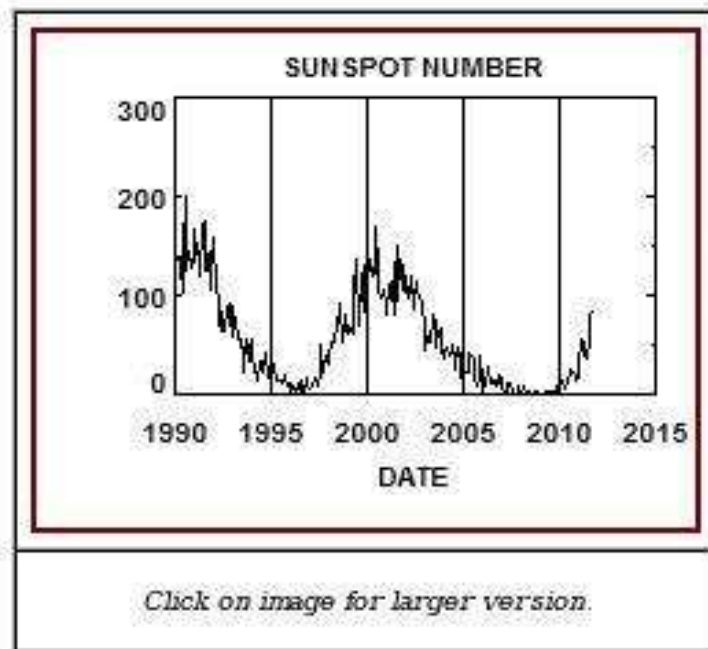
- Slné škvvrny:



<http://www.dailymail.co.uk/sciencetech/article-2042428/Best-auroras-seen-Britain-thanks-huge-solar-flares.html>

Motivácia

- Slné škvvrny:



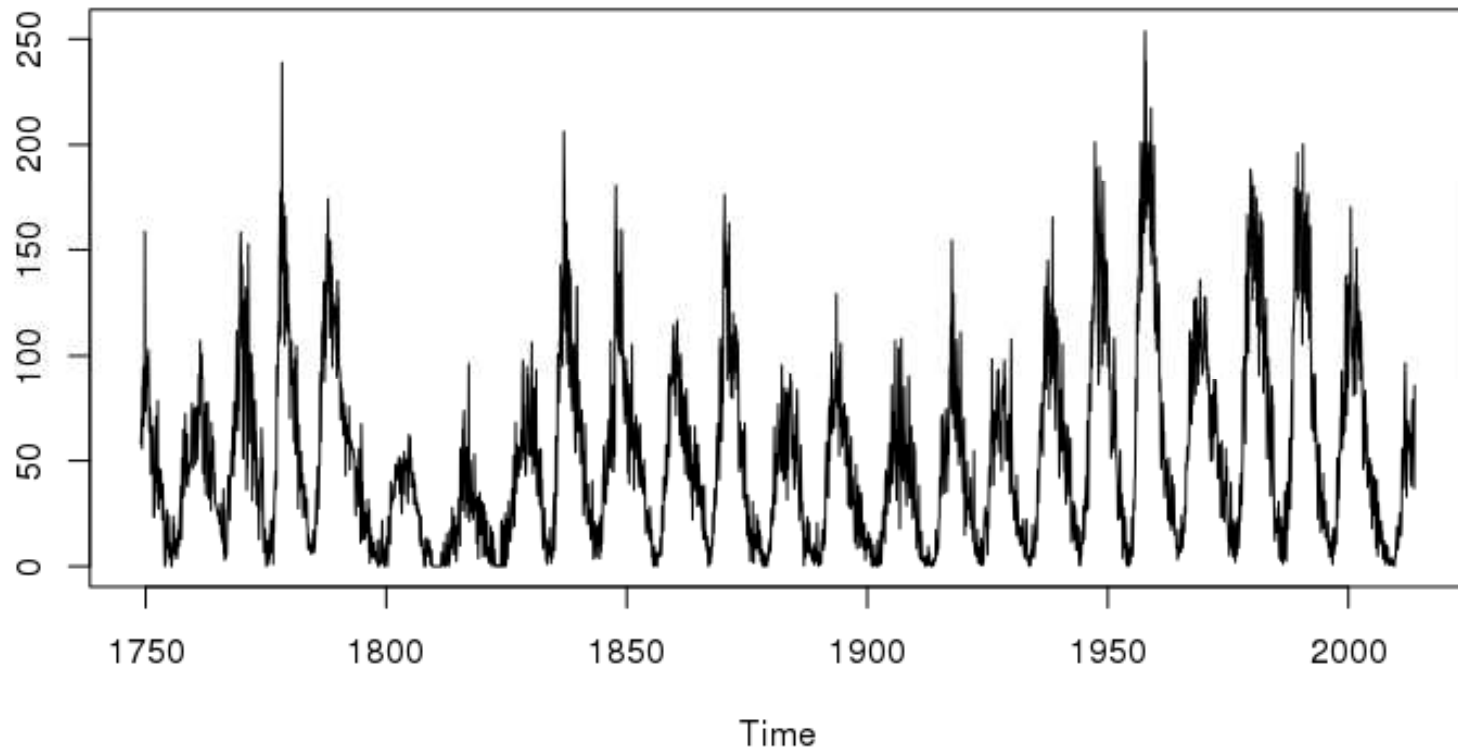
Sunspot Numbers

In 1610, shortly after viewing the sun with his new telescope, Galileo Galilei (or was it Thomas Harriot?) made the first European observations of Sunspots. Continuous daily observations were started at the Zurich Observatory in 1849 and earlier observations have been used to extend the records back to 1610. The sunspot number is calculated by first counting the number of sunspot groups and then the number of individual sunspots.

<http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

Motivácia

- Slnečné škvrny - graf z dlhšieho obdobia:



Dáta zo stránky <http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

- Otázka: Ako určiť periódu?

Spektrum

- Postupnosť $\{\gamma_j\}_{j=-\infty}^{\infty} \rightarrow$ vytvárajúca funkcia

$$g(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j$$

- Stacionárny proces Y s autokovarianciami $\{\gamma_j\}_{j=-\infty}^{\infty} \rightarrow$ spektrum

$$s_Y(\omega) = \frac{1}{2\pi} g(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

kde i je imaginárna jednotka.

Vlastnosti spektra

- Spektrum $s_Y(\omega)$:
 - ◇ dá sa napísať v tvare:

$$s_Y(\omega) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(\omega j) \right]$$

- ◇ nadobúda reálne hodnoty
 - ◇ je párna funkcia
 - ◇ je periodická funkcia s periódou 2π
- stačí nám poznať hodnoty spektra medzi 0 a π
- Dá sa dokázať, že $s_Y(\omega) \geq 0$ [Fuller, 1976]

Spektrum a disperzia

- Výpočet autokovariancií zo spektra:

$$\gamma_k = \int_{-\pi}^{\pi} s_Y(\omega) e^{i\omega k} d\omega$$

- Pre $k = 0$ dostaneme: $\gamma_0 = \int_{-\pi}^{\pi} s_Y(\omega) d\omega$, a keď že spektrum je párna funkcia:

$$\gamma_0 = 2 \int_0^{\pi} s_Y(\omega) d\omega,$$

čiže disperzia γ_0 je dvojnásobkom plochy pod spektrom na intervale $[0, \pi]$

- Z priebehu spektra teda vidíme, ktoré frekvencie najviac prispievajú k disperzii procesu - tie, kde má spektrum vysokú funkčnú hodnotu

Odhad spektra - výberový periodogram

- Máme dáta $y_1, \dots, y_T \rightarrow$ chceme z nich odhadnúť spektrum časového radu
- Prvá myšlienka: nahradíme autokovariancie v definícii spektra výberovými autokovarianciami, takto dostaneme - výberový periodogram:

$$\hat{s}_y(\omega) = \frac{1}{2\pi} \sum_{j=-T+1}^{T-1} \hat{\gamma}_j e^{-i\omega j} = \frac{1}{2\pi} \left[\hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} \hat{\gamma}_j \cos(\omega j) \right]$$

- Nevýhody:
 - ◇ odhady majú veľkú disperziu
 - ◇ presnosť sa nezlepšuje so zvyšujúcim sa počtom dát (lebo odhadujeme stále viac autokovariancií)
- \rightarrow potrebujeme iný odhad spektra

Odhad spektra - výberový periodogram

- Nevýhody - podrobnejšie:
 - ◇ [Fuller, 1976]: pre veľký rozsah výberu má podiel $\frac{2\hat{s}_y(\omega)}{s_Y(\omega)}$ približne $\chi^2(2)$ rozdelenie a tieto podiely sú pre rôzne ω približne nezávislé
 - ◇ $E[\chi^2(2)] = 2$, a teda

$$E[\hat{s}_y(\omega)] \sim s_Y(\omega)$$

- to je OK

- ◇ Ale 95 percentný interval spoľahlivosti pre $\chi^2(2)$ je $(0.05, 7.4)$ a teda IS pre spektrum je

$$(0.025\hat{s}_y(\omega), 3.7\hat{s}_y(\omega))$$

- príliš široký

Odhad spektra - vylepšenie

- Idea: ak sú blízke hodnoty frekvencií, majú podobné hodnoty spektra → za odhad spektra pre danú frekvenciu zoberieme vážený **priemer hodnôt výberového periodogramu \hat{s}_y** z okolitých frekvencií:

$$(1) \quad \hat{s}_Y(\omega_j) = \sum_{m=-h}^h \kappa(\omega_{j+m}, \omega_j) \hat{s}_y(\omega_{j+m})$$

kde

- ◇ $\omega_j = 2\pi j/T$
- ◇ konštanta h udáva počet okolitých frekvencií, ktoré berieme do úvahy pri výpočte odhadu (nazýva sa **bandwidth**)
- ◇ funkcia κ určuje **váhy**, s ktorými tieto frekvencie vstupujú do odhadu (v súčte dávajú 1)

Príklad 1: simulované dáta a postup v R

- Pripomeňme si AR(2) proces s periodickým charakterom:

$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$$

- Vieme:

- ◇ korelácie splňajú diferenčnú rovnicu

$$\rho(t) - 1.4\rho(t-1) + 0.85\rho(t-2) = 0,$$

ktorá má všeobecné riešenie

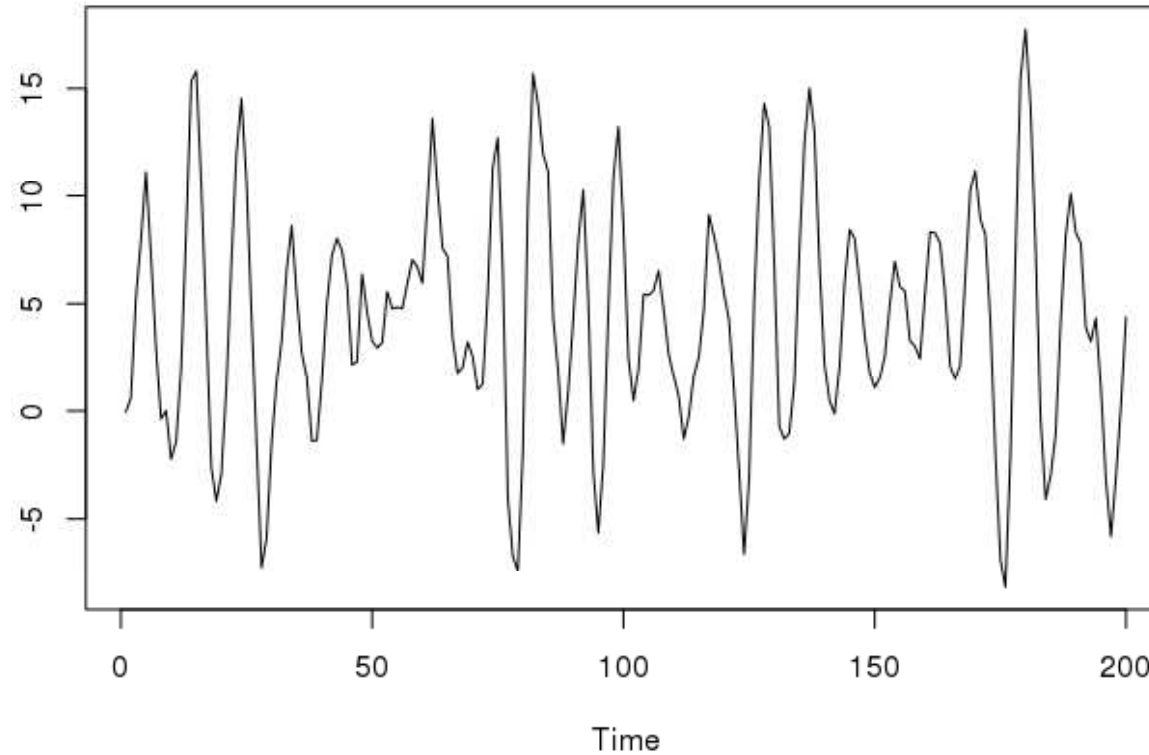
$$\rho(t) = 0.922^t (c_1 \cos(0.709t) + c_2 \sin(0.709t))$$

- ◇ sínus a kosínus vo všeobecnom riešení:

$$\cos(kt), \sin(kt) \rightarrow \text{perióda } \frac{2\pi}{k} = 8.862 \approx 9$$

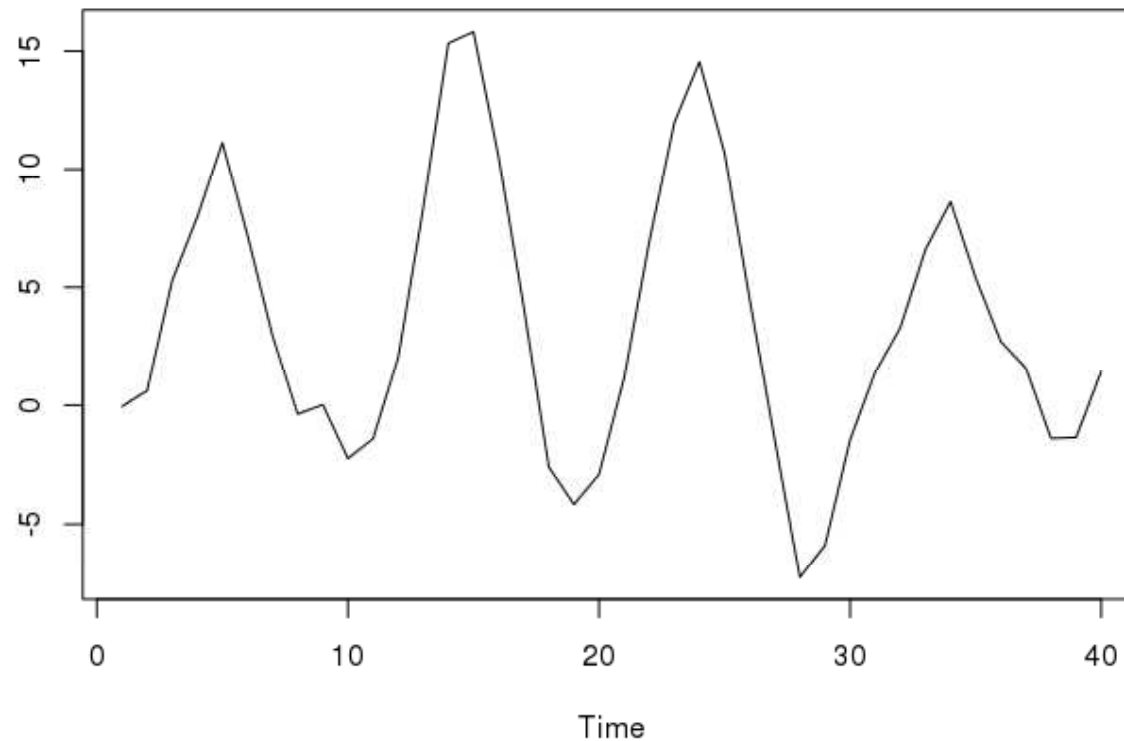
Príklad 1: simulované dáta a postup v R

- Vygenerovaný priebeh:



Príklad 1: simulované dáta a postup v R

- Vygenerovaný priebeh - kratší časový interval (lepšie vidíme periódu):

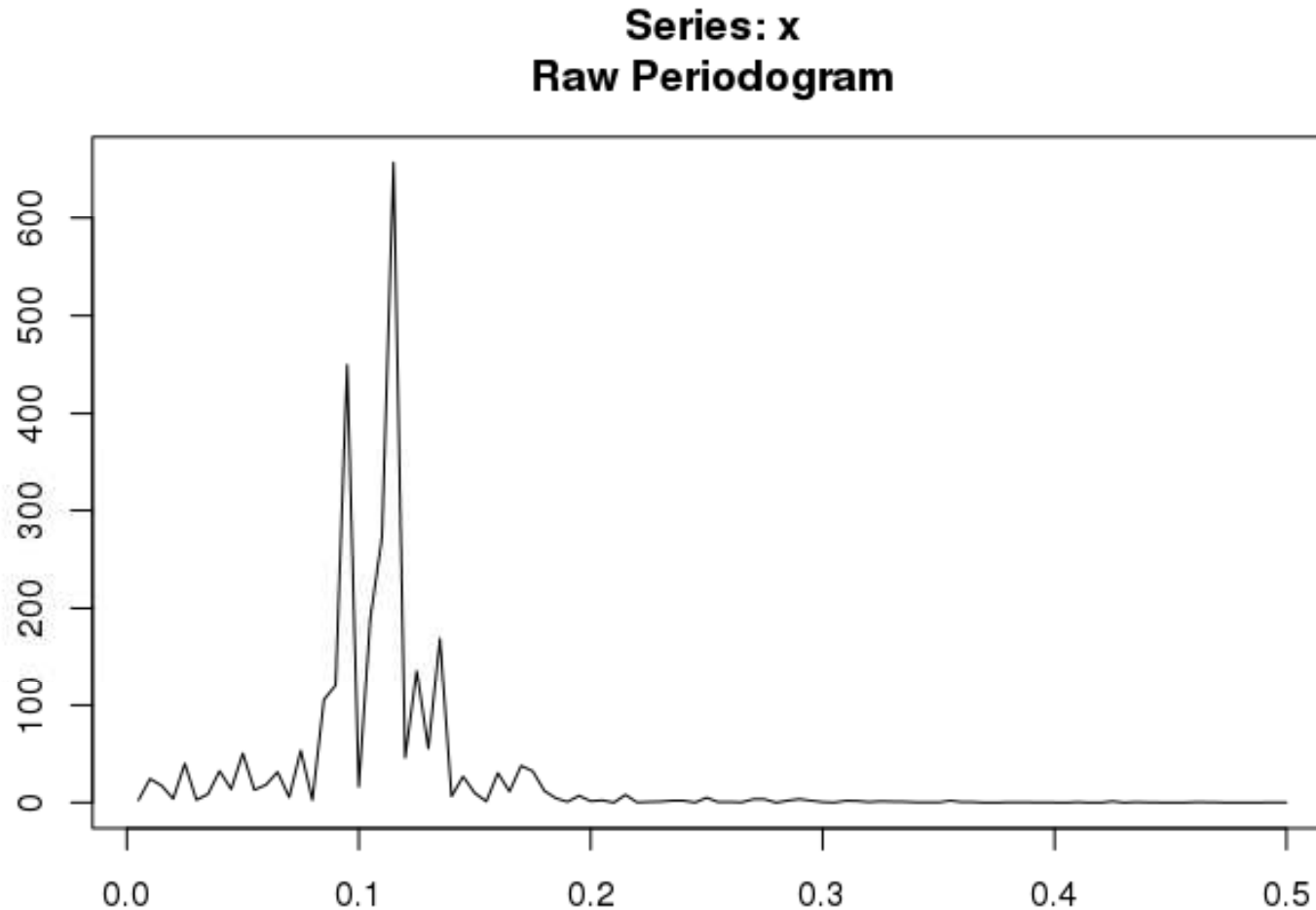


Príklad 1: simulované dáta a postup v R

- Odhad spektra v R-ku:
 - ◇ výberový periodogram:
`spectrum(x, log="no")`
 - ◇ zhladený výberový periodogram:
`spectrum(x, kernel("daniell"), log="no")`
`spectrum(x, kernel("modified.daniell"), log="no")`
 - ◇ iné škálovanie x-ovej osi: od 0 do 1/2 (namiesto do $\pi/2$); takto: **perióda = 1/frekvencia**)

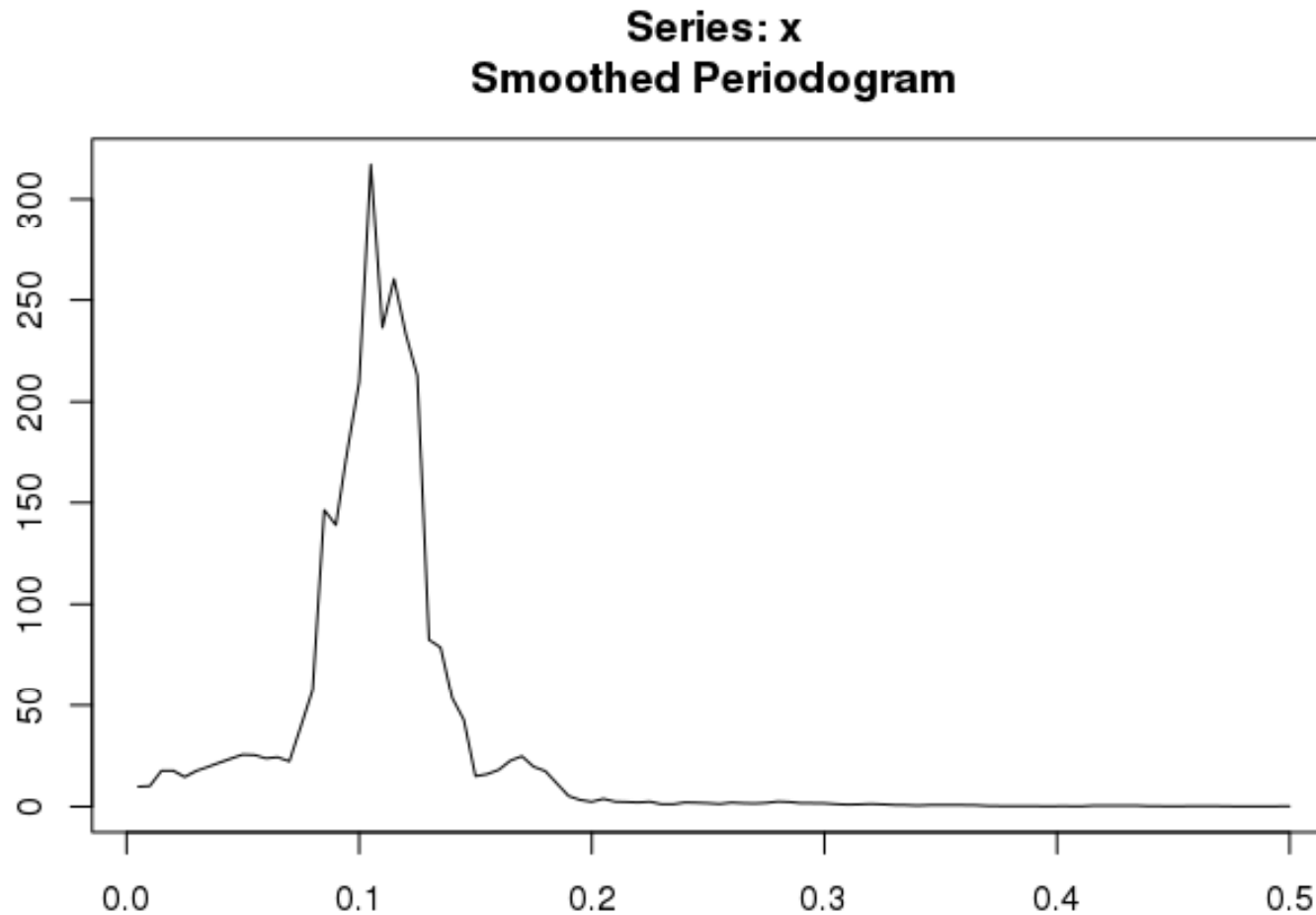
Príklad 1: simulované dáta a postup v R

- Výberový periodogram: `spectrum(x, log="no")`



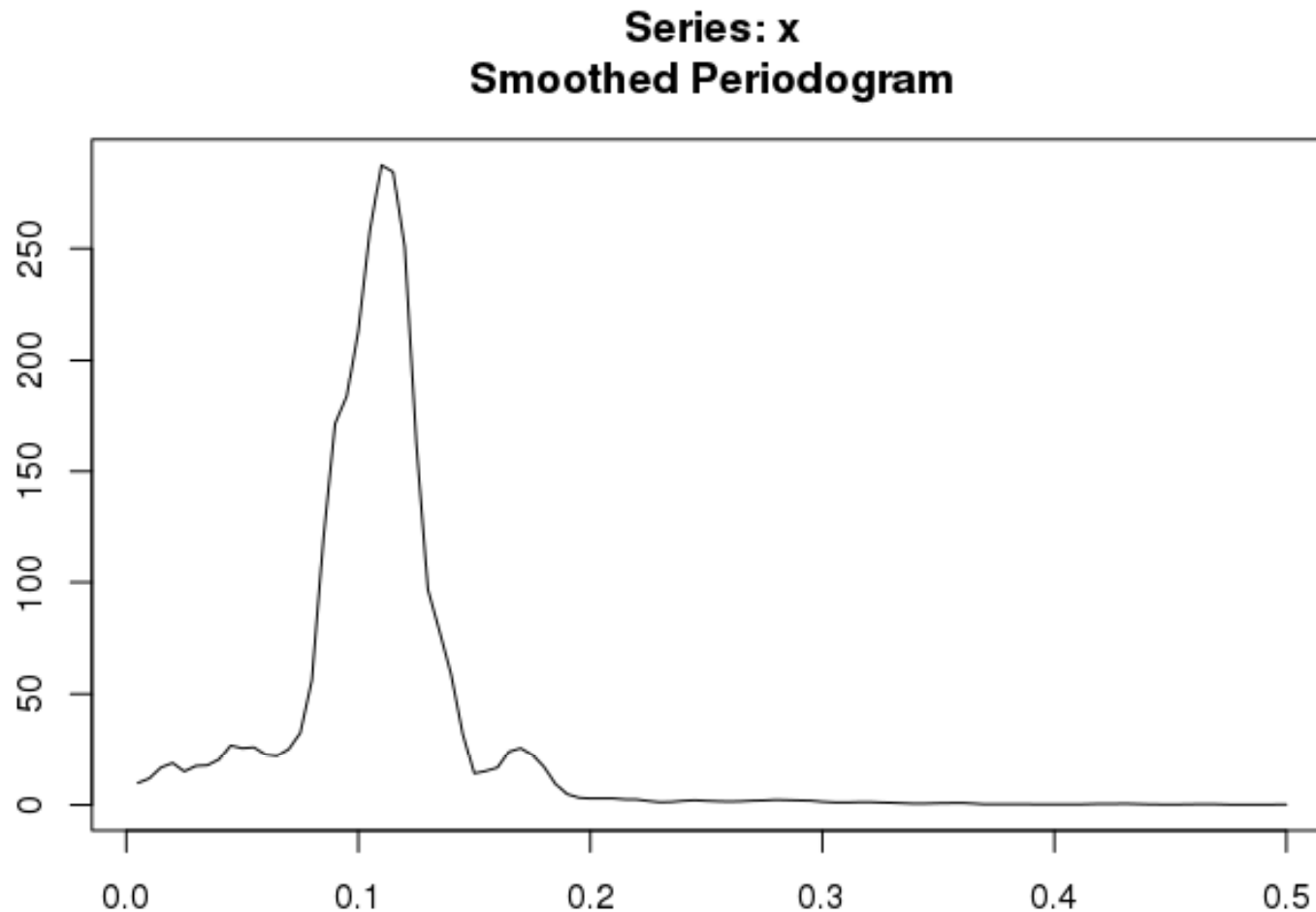
Príklad 1: simulované dáta a postup v R

- Zhladený periodogram:
`spectrum(x, kernel("daniell"), log="no")`



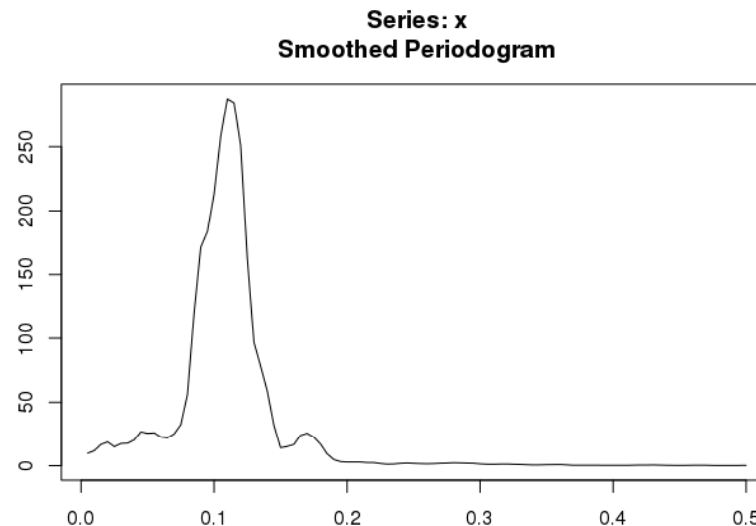
Príklad 1: simulované dáta a postup v R

- Zhladený periodogram:
`spectrum(x, kernel("modified.daniell"), log="no")`



Príklad 1: simulované dáta a postup v R

- Pripomeňme si:



- Nájďeme maximum a zodpovedajúcu periódu:

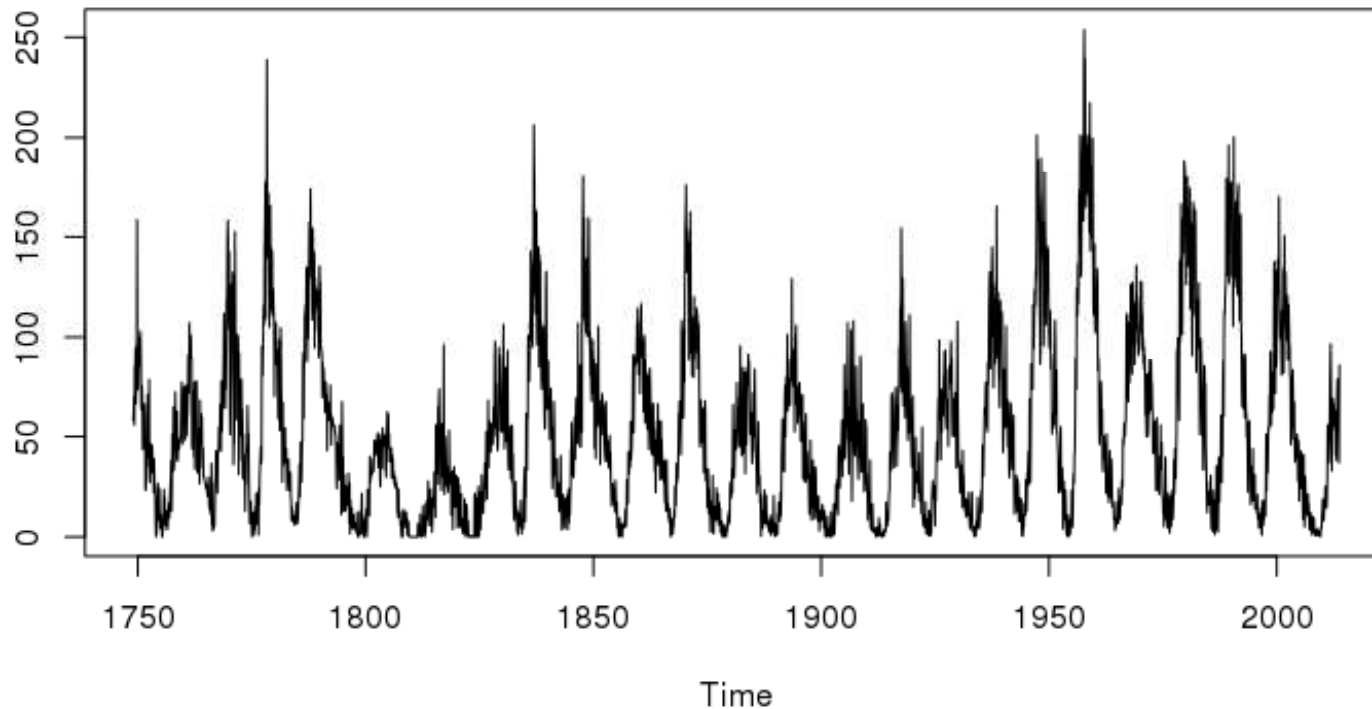
```
> sp=spectrum(x, kernel("modified.daniell"),log="no")
> max(ss$spec)
[1] 287.7513
> sp$freq[which.max(sp$spec)]
[1] 0.11
>
> 1/sp$freq[which.max(sp$spec)]
[1] 9.090909
```

Príklad 2: slnečné škvrny

- Priebeh dát:

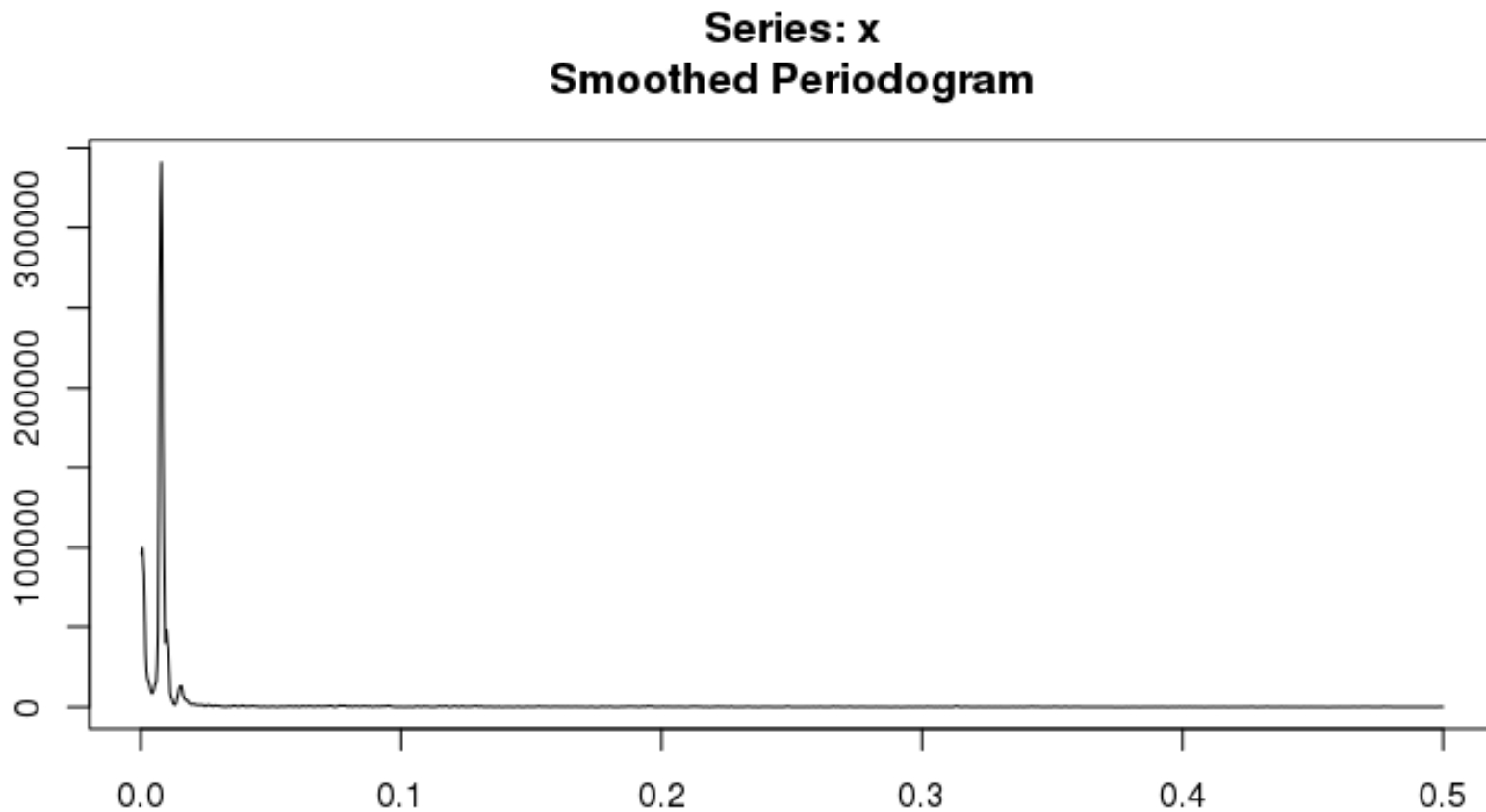
```
sun=read.table("sun.txt")
```

```
plot(ts(sun),frequency=12,start=c(1749,1))
```



Príklad 2: slnečné škvrny

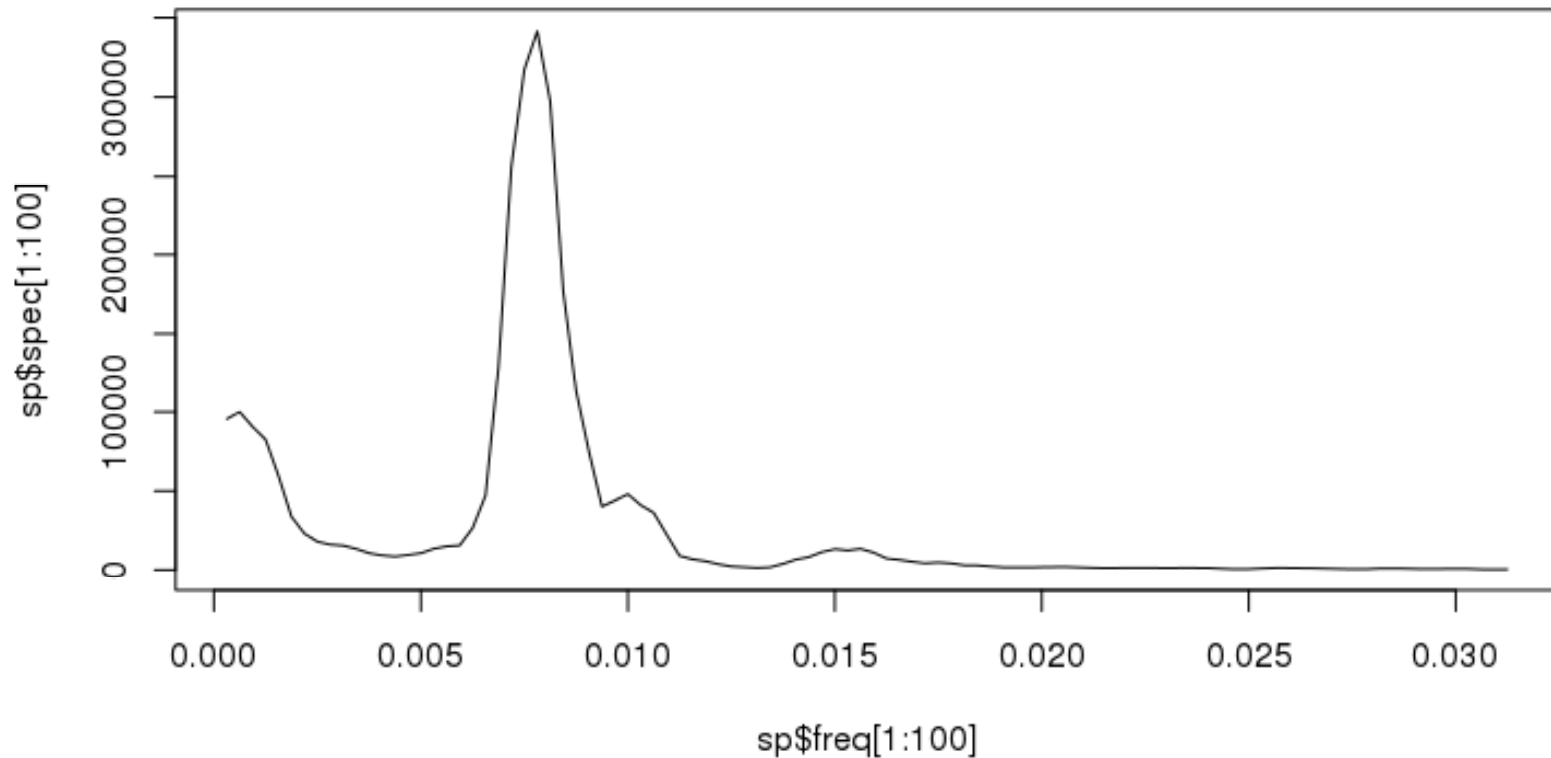
- Odhad spektra:
`spectrum(sun, kernel("modified.daniell"), log="no")`



Príklad 2: slnečné škvrny

- Odhad spektra - výrez:

```
plot(sp$freq[1:100], sp$spec[1:100], type="l")
```



Príklad 2: slnečné škvrny

- Nájďeme maximum odhadnutého spektra a periódu:

```
> 1/sp$freq[which.max(sp$spec)]  
[1] 128  
> 1/sp$freq[which.max(sp$spec)]/12  
[1] 10.66667
```

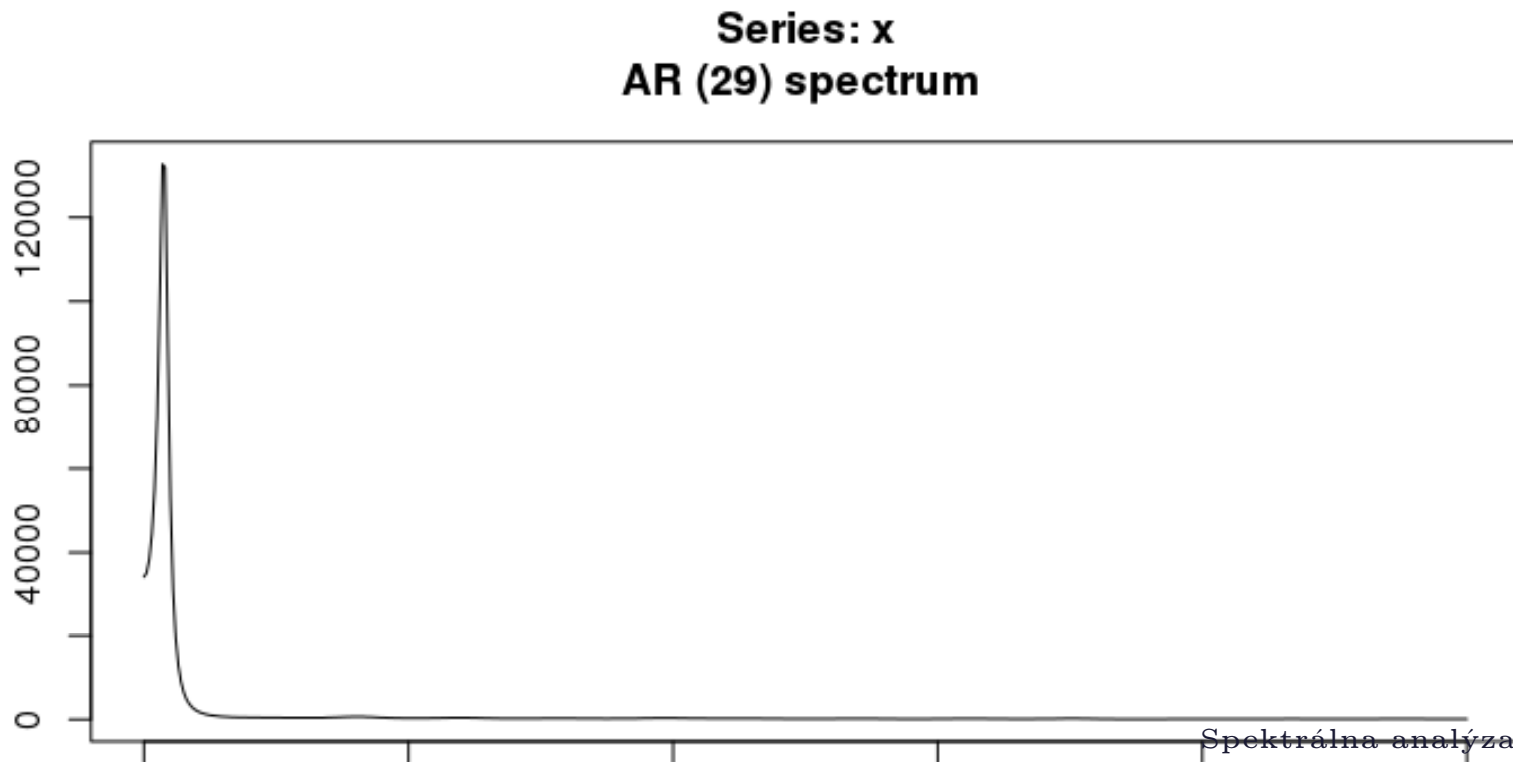
- Text zo stránky:

The "sunspot number" is then given by the sum of the number of individual sunspots and ten times the number of groups. Since most sunspot groups have, on average, about ten spots, this formula for counting sunspots gives reliable numbers even when the observing conditions are less than ideal and small spots are hard to see. Monthly averages (updated monthly) of the sunspot numbers ([181 kb JPEG image](#)), ([307 kb pdf-file](#)), ([62 kb text file](#)) show that the number of sunspots visible on the sun waxes and wanes with an approximate 11-year cycle.

<http://solarscience.msfc.nasa.gov/SunspotCycle.shtml>

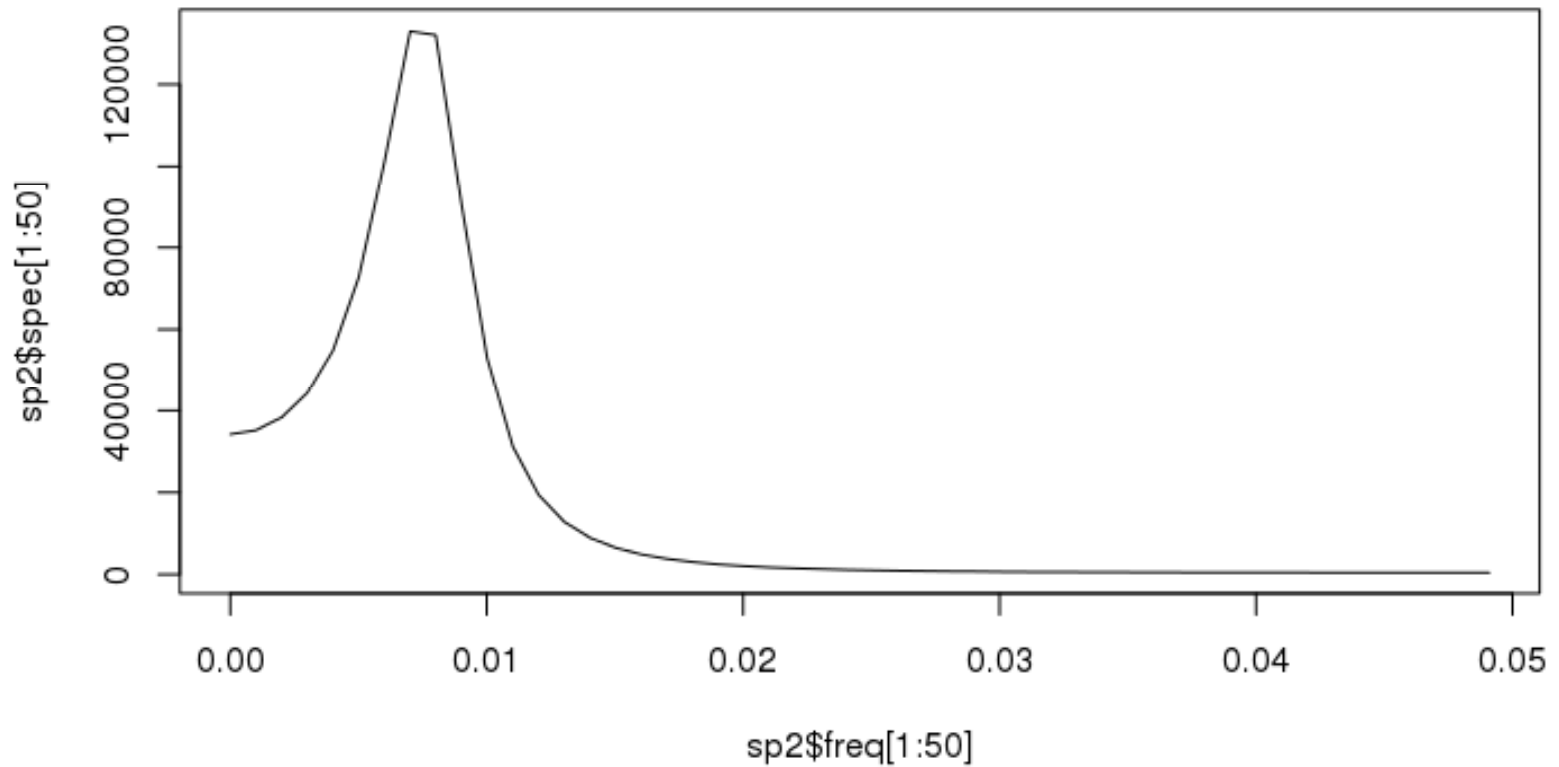
Príklad 3: odhad spektra AR procesmi

- Dáta modelujeme AR procesom
- Spektrum odhadneme ako spektrum získaného AR procesu
- V R-ku pre slnečné škvrny: `sp2=spectrum(ts(sun), method="ar", log="no")`



Príklad 3: odhad spektra AR procesmi

- Výrez:



Príklad 3: odhad spektra AR procesmi

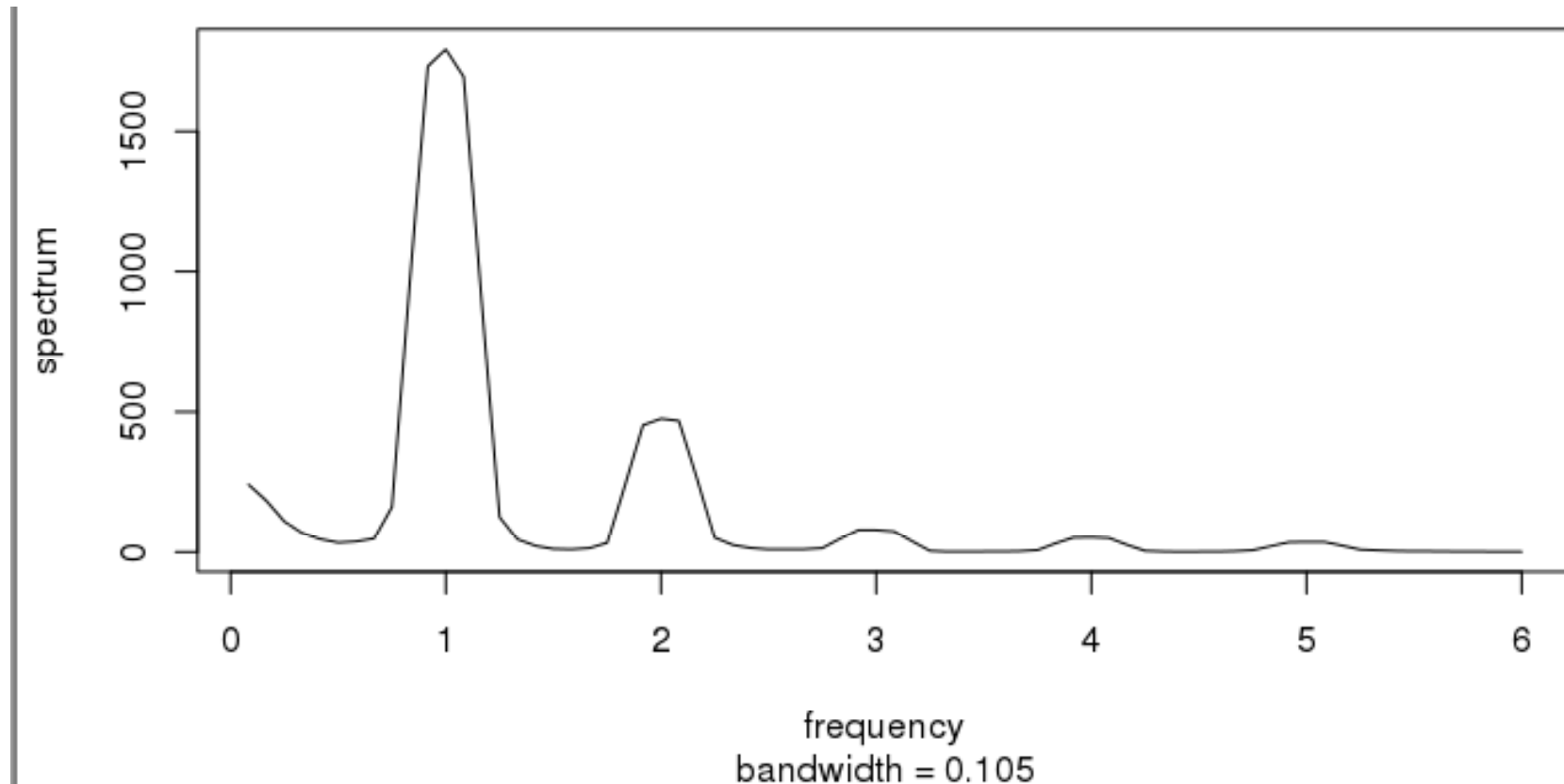
- Nájdenie maxima:

```
> sp2=spectrum(ts(sun),method="ar",log="no")
>
> 1/sp2$freq[which.max(sp2$spec)]
[1] 142.5714
> 1/sp2$freq[which.max(sp2$spec)]/12
[1] 11.88095
```

Poznámky k R-ku

- Pri prvom postupe (zhladený výberový periodogram):
 - ◇ Z dát sa odstráni lineárny trend.
 - ◇ Ak majú dáta časovú štruktúru, zachová pri výpočte frekvencie a periódy.
- PRÍKLAD: cestujúci aerolinkami
 - ◇ v R-ku:
`data(AirPassengers)`
`x=AirPassengers`
 - ◇ takto načítané dáta už majú časovú štruktúru:
jednotka času = rok, frekvencia dát = 12

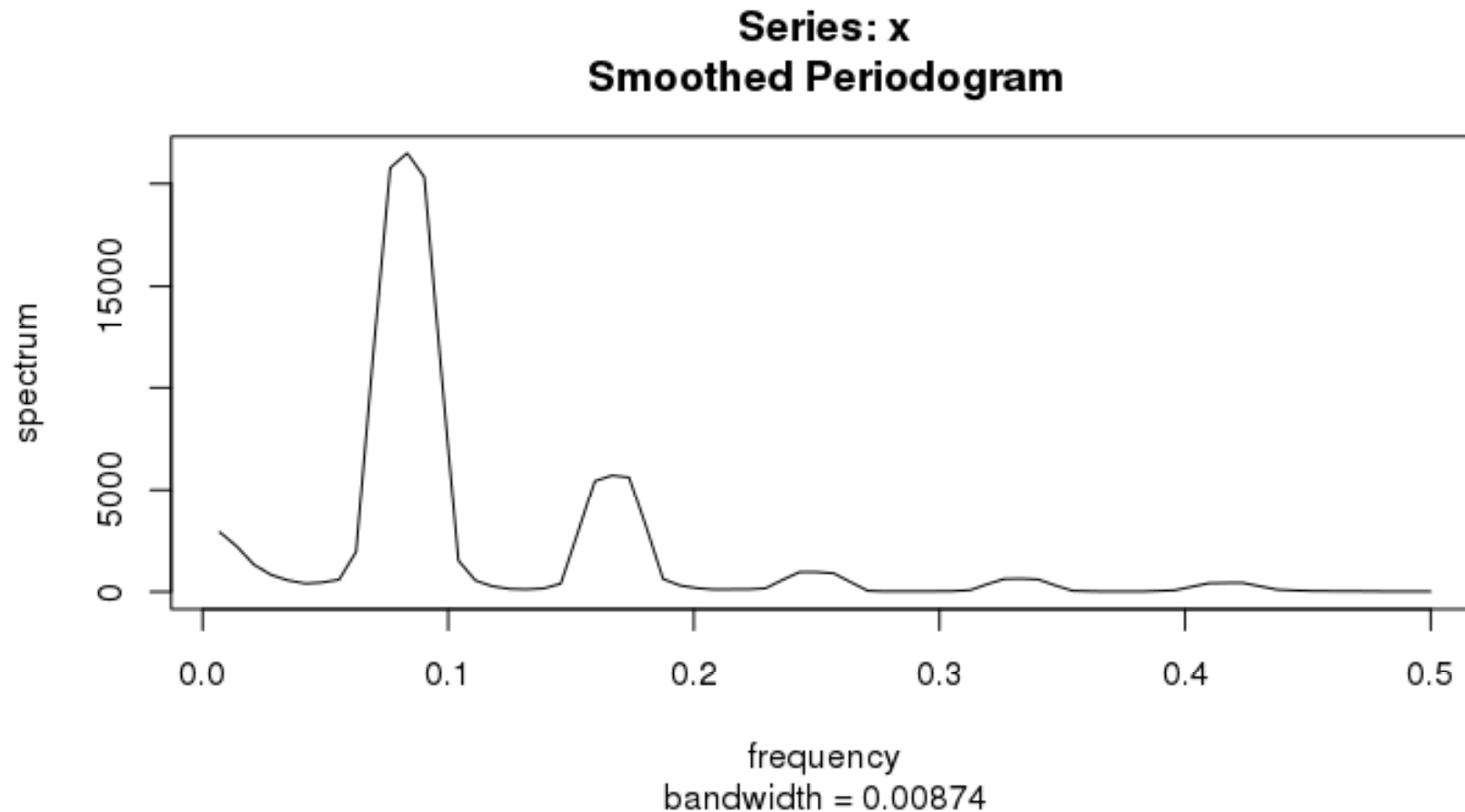
Poznámky k R-ku



```
> data(AirPassengers)
> x=AirPassengers
> sp=spectrum(x, kernel("modified.daniell"), log="no")
> 1/sp$freq[which.max(sp$spec)]
[1] 1
```

Poznámky k R-ku

Pre porovnanie, bez špeciálnej časovej štruktúry:



```
> sp=spectrum(ts(x, frequency=1),kernel("modified.daniell"),log="no")
> 1/sp$freq[which.max(sp$spec)]
[1] 12
```