

# Czy / jak R może zastąpić Excela w firmie?

Artur Suchwałko

QuantUp ([quantup.pl](http://quantup.pl)), MedicWave ([spectrolyzer.com](http://spectrolyzer.com))

2014.10.16

Wstę

# Kontekst: praktyczna analiza danych w biznesie

- R vs. Excel – porównanie i trudności, które mogą być powodem do zmiany
- Jak często wygląda praca analityczna w firmach?
- Gdzie R może zastąpić Excela?

# Porównanie gruszek z jabłkami?

- Wydaje się, że R i Excel nie można ich porównywać, bo to narzędzia służące do (trochę) innych celów
- Ale... Excel jest często jedynym narzędziem analitycznym i służy do wszystkiego

# Teoria a praktyka

- Bardzo wiele prac analitycznych (większość!) jest wykonywanych poza narzędziami statystycznymi
- Na przykład w Excelu

# O jakich firmach i na bazie jakich doświadczeń?

- Raczej o korporacjach
- Między innymi o instytucjach finansowych
- Doświadczenie z pracy wewnątrz organizacji
- Doświadczenie konsultanta
- Nie mówimy tutaj o typowym, zaawansowanym, modelowaniu statystycznym

R vs. Excel

# Dlaczego R?

- Nie trzeba (na ogół) nikogo przekonywać
- Zwłaszcza w takiej grupie
- Elastyczność
- Powtarzalność analiz
- Nic nie kosztuje
- Wielość dostępnych metod i zastosowań
- Wygoda użycia i programowania
- Szybkość
- Popularność – standard!



# Dlaczego jednak Excel?

- „Wspólny język” w firmach – trudno zmienić standard
- Użycie Excela jest często wygodniejsze
- Praca w Excelu jest czasem wygodniejsza (bo interaktywna, szybka i wszystko widać)
- Excel zawsze jest dostępny w firmie
- Dobrze pracuje z innymi narzędziami MS Office
- Łatwiej zaufać wynikowi przygotowanemu w Excelu, jeśli widać wszystkie kroki analizy

Problemy...

# Problemy z Excelem

- Wspiera niechlujstwo (wklejanie wartości!)
- Daje wynik natychmiast – nie trzeba dwa razy pomyśleć
- Łatwość korekt i samodzielnych zmian w formułach i rozwiązaniach
- Trzeba ostrożnie aktualizować dane, np. przesuwanie formuł
- To są niekoniecznie problemy tego konkretnego narzędzia ;-)
- Szczegółowe problemy wymienione na blogu Patricka Burnsa: → *Spreadsheet Addiction*

# Rozwiązania niektórych z tych problemów

- Dobre praktyki w pracy z Excelem
- Świadomość problemów i ich unikanie

# Problemy z R w środowisku biznesowym – ludzie

- W R zawsze trzeba programować
- Trzeba przestawić się na inny, R-owy, sposób myślenia o analizie danych
- Brak wykwalifikowanych ludzi w firmie
- Trzeba pracownikom dać czas na doształcenie – na co dzień
- być może zmniejszając liczbę bieżących zadań
- Konieczność inwestycji czasu w naukę i programowanie analiz (stroma krzywa uczenia R, a analitycy nie mają czasu załadować!)
- Złe kody
- Bardzo złe kody, np. setki linii pisanych we wbudowanym edytorze R

# Problemy z R w środowisku biznesowym – organizacja

- R często jest używany w pewnych departamentach, ale rzadko jest narzędziem z wyboru dla całej firmy
- Trzeba promować R w firmie, jeśli stanie się narzędziem z wyboru, to jest łatwiej
- Konieczność utrzymania wykwalifikowanych ludzi → porzucone kody (ale zdarzają się też porzucone makra VB)
- Konieczna jest wygodna komunikacja z innymi, którzy nie znają R
- Wymaganie formatu np. Excela oraz dostępu do danych i formuł (metod analizy) dla wszystkich
- Wymaganie edytowalnych wykresów

# Problemy z R w środowisku biznesowym – technika

- Opór w stosunku do „nowych technologii”
- Brak uświadomionych głębszych potrzeb analitycznych
- Problemy z użyciem R, np.:
  - ▶ względy bezpieczeństwa
  - ▶ 32-bitowe systemy operacyjne
  - ▶ słabe komputery

# Rozwiązania niektórych z tych problemów

- Często są bardzo trudne i wymagają zmiany sposobu myślenia o analizie danych
- Wymagają też zmiany w doborze analityków
- oraz utrzymywania wykwalifikowanych analityków
- Wymagają zmiany procesów i organizacji pracy



Jak (często) wygląda praca analityczna w firmach?

# Jak (często) wygląda praca analityczna w firmach? – rodzaje zadań

- Zadania ad hoc – „końcówka raportująca”
- Zadania powtarzalne
- ale za każdym razem (trochę) inne
- Wiele zbędnych raportów, których odbiorcy nie używają, ale wymagają (bo zawsze były)
- Im szybciej tym prędzej!

# Jak (często) wygląda praca analityczna w firmach? – analiza

- Wklejanie / klejenie wyniku
- Łączy między wieloma plikami
- VB może wszystko!
- Mimo posiadania hurtowni danych sporo przetwarzania danych wykonywanego jest na danych eksportowanych do Excela
- lub na danych wyeksportowanych z Excela (czasami to jest wykonywane w R)
- Budowa rozwiązań analitycznych na bazie R jest rzadka
- Szczególnie rzadka jest budowa porządných takich narzędzi analitycznych

# Jak (często) wygląda praca analityczna w firmach? – metody

- Trend liniowy
- Wychodzi tyle, ile ma wyjść
- Korekty eksperckie / zarządcze
- Tyle samo, ile było w zeszłym roku
- O 10% więcej niż w zeszłym roku
- Rozbijmy plan sumaryczny na poszczególne sklepy – i mamy prognozę

# Jakie zadania najczęściej realizowane są w R?

- Oczyszczanie i przetwarzanie danych, np. tekstowych i dat – coraz częściej w R zamiast po stronie bazy danych czy w Excelu
- Proste modelowanie
- Powtarzalne modelowanie i przetwarzanie danych
- Proste aplikacje analityczne z wykorzystaniem Shiny

# W jakich zadaniach R może zastąpić Excela?

- Cykliczne oczyszczanie i przetwarzanie danych
- Wizualizacja dla celów eksploracji danych
- Tabele przestawne i podobne operacje – szybkość i wygoda!
- Automatyzacja pracy, np. dopasowywanie prostych modeli w segmentach
- Prognozowanie
- Udostępnianie wyników prac analitycznych w postaci narzędzi – Shiny

# Grafika i statystyka

- Nie zajmujemy się grafiką
- Nie zajmujemy się zaawansowaną statystyką

# Typowe funkcje arkusza

- Na blogu Patricka Burnsa jest długa lista funkcji Excela z tłumaczeniem ich na polecenia R: → *From spreadsheet thinking to R thinking*



# Oczyszczanie i przetwarzanie danych

- Oczyszczanie i przetwarzanie danych, np. pakiety `stringr`, `dplyr`, `data.table`, `sqldf`, `reshape2`

# Różne, bardziej zaawansowane

- Szukaj wyniku (rozwiązywanie równań nieliniowych): optymalizacja, np. `uniroot`, link na końcu prezentacji
- Tabele przestawne (grupowanie danych): funkcja `aggregate`, pakiet `reshape2`, pakiet `data.table`
- Solver (zadanie programowania liniowego): pakiet `linprog`, funkcja `solveLP`

# Excel Analysis Toolpak

→ *Excel Analysis Toolpak*

- Statystyki opisowe: wiadomo
- Histogram: funkcja `hist`
- Generowanie liczb z różnych rozkładów: np. `*r*norm`, `*d*norm`, `*p*norm`, `*q*norm`
- Próbkowanie: funkcja `sample`
- Średnia ruchoma i wygładzanie wykładnicze: pakiet `forecast` i funkcje `ma` oraz `ses`
- Testy statystyczne: np. `t-test`, resztę łatwo znaleźć
- Analiza wariancji: funkcja `aoV`
- Regresja liniowa: funkcja `lm`

Porcja informacji dla „przechodzących na  
stronę R”

# Jak przejść z Excela na R

- Dlaczego Excel jest zły: → *Spreadsheet Addiction*
- Zmiana sposobu myślenia z excelowego na R-owe + wielka lista funkcji: → *From spreadsheet thinking to R thinking*
- Przejście z Excela na R: → *A first step towards R from spreadsheets*
- Książka: „R for Dummies”, A. de Vries, J. Meys, rozdział 19: Ten Things You Can Do in R That You Would've Done in Microsoft Excel
- R dla użytkowników Excela – krótkie video: → *A video introduction to R for Excel users*
- Nauka R dla znających Excela: → *Learn R Toolkit To Help Excel Users Move Up To R*
- Kiedy używać Excela, a kiedy R: → *When to use Excel, when to use R?*

# Dodatkowe informacje – konkretne przepisy

Seria wpisów na blogu QuantUp:

- → *Linia trendu z MS Excel w R*
- → *Funkcja „szukaj wyniku” (“goal seek”) z MS Excel w R*
- → *Użycie pakietów `dplyr` i `data.table` w pracy z dużymi zbiorami danych*
- → *Wykresy słupkowe (kolumnowe) w R i w Excelu*

Nie trzeba wybierać – te dwa narzędzia można połączyć

→ *A million ways to connect R and Excel*

# Podsumowanie



# Jak to w końcu jest?

- Ani „TAK” ani „NIE” ;-)
- Zmiana Excela na R to nie jest wyłącznie kwestia porównania narzędzi
- ani tego, jakie zadanie w jakim narzędziu lepiej / łatwiej wykonać.
- Zależy to od specyfiki pracy, środowiska, departamentu i przełożonych
- oraz „kultury analitycznej”
- Wyjątkowo mocno zależy od bezpośredniego przełożonego

# Prezentacja

... jest na blogu QuantUp: → [quantup.pl/blog/](http://quantup.pl/blog/)

# Kontakt?

Jeśli potrzebujesz, to napisz maila: artur *[at]* quantup *[dot]* pl

# Dyskusja

Co o tym sądzicie?