

# Analiza fanfiction w fandomie serii *Harry Potter* w latach 2004-2019

Barbara Kulawska

AGH, Wydział Informatyki, Elektroniki i Telekomunikacji  
*Rachunek prawdopodobieństwa i statystyka 2020/2021*

Kraków, 22 stycznia 2021

## Spis treści

<b>1</b>	<b>Streszczenie raportu</b>	<b>2</b>
<b>2</b>	<b>Opis danych</b>	<b>2</b>
<b>3</b>	<b>Cele projektu</b>	<b>2</b>
<b>4</b>	<b>Analiza cech ilościowych</b>	<b>3</b>
4.1	Liczba rozdziałów . . . . .	3
4.2	Liczba słów . . . . .	5
4.3	Liczba dodań do ulubionych . . . . .	7
<b>5</b>	<b>Analiza cech jakościowych</b>	<b>9</b>
5.1	Gatunek . . . . .	9
5.2	Język . . . . .	10
5.3	Rating . . . . .	10
5.4	Główny bohater . . . . .	11
5.5	Główny paring . . . . .	12
5.6	Data publikacji . . . . .	13
<b>6</b>	<b>Analiza zależności między danymi</b>	<b>14</b>
6.1	W jaki sposób liczba rozdziałów zależy od liczby słów opowiadania? W jaki sposób liczba dodań do ulubionych zależy od liczby słów opowiadania? . . . . .	14
6.2	Jak różnią się najpopularniejsi bohaterowie w zależności od języka? . . . . .	16
6.3	Jak różnią się najpopularniejsze paringi w zależności od języka? . . . . .	17
<b>7</b>	<b>Analiza wpływu czynników zewnętrznych na dane</b>	<b>20</b>
<b>8</b>	<b>Wnioski</b>	<b>20</b>

# 1 Streszczenie raportu

Raport powstał w oparciu o analizę danych dotyczących literackiej twórczości fanowskiej w fandomie *Harry'ego Pottera*.

Analizy dokonano przy użyciu języka R.

Celem analizy była eksploracja danych w taki sposób, by móc wyznaczyć kierunki i sposoby kształtowania się trendów wśród fanów. Projekt łączy praktyczne wykorzystanie wiedzy na temat analizy danych, którą zdobyłam podczas zajęć z przedmiotu *Rachunek prawdopodobieństwa i statystyka*, z hobbystycznym charakterem zagadnienia, którym interesuję się od wielu lat.

Raport zawiera dokładny opis danych, cele projektu (pytania i problemy, którymi się zajmowano), analizę cech ilościowych i jakościowych oraz zależności między nimi, a także omówienie wpływu czynników zewnętrznych na kształtowanie się danych.

## 2 Opis danych

Dane do projektu pochodzą ze strony: [www.kaggle.com/nehatiwari03/harry-potter-fanfiction-data](http://www.kaggle.com/nehatiwari03/harry-potter-fanfiction-data).

Dotyczą one zagadnienia twórczości fanowskiej (tzw. **fanfiction**<sup>1</sup>), stworzonej przez fanów w oparciu o serię książek *Harry Potter* autorstwa J.K. Rowling. Zestaw danych zawiera informacje o wszystkich takich tekstach opublikowanych na portalu **fanfiction.net** w latach 2004-2019.

Dla każdego z opowiadań zebrano następujące dane: liczba rozdziałów (**chapters**), liczba słów (**words**), rok opublikowania (**published**), gatunek literacki (**genre**), język (**language**), ograniczenie wiekowe (**rating**), liczba użytkowników portalu, którzy dodali historię do ulubionych (**favs**), głównego bohatera (**main.character**) oraz główny paring<sup>2</sup> (**pairing**).

Dane możemy podzielić na dwie kategorie: cechy ilościowe (chapters, words, favs) oraz cechy jakościowe (pozostałe).

## 3 Cele projektu

Na podstawie analizy danych, którą przeprowadzę, postaram się odpowiedzieć na następujące pytania:

- Jak wygląda statystyka liczby rozdziałów, słów i dodań do ulubionych dla opowiadań? Jakie zależności można zaobserwować między nimi?
- Jaki gatunek literacki dominuje w twórczości fanowskiej?
- W jakim języku powstaje najwięcej opowiadań?
- Jaki jest rozkład ograniczeń wiekowych w opowiadaniach?

---

<sup>1</sup>Fanfiction – dzieła literackie tworzone nieoficjalnie przez fanów danego utworu (filmu, książki, serialu itp.), wykorzystujące postacie i świat z tego utworu. Więcej: [pl.wikipedia.org/wiki/Fan\\_fiction](http://pl.wikipedia.org/wiki/Fan_fiction)

<sup>2</sup>Paring (pairing) – określenie dwóch postaci ze świata fikcyjnego, które są ze sobą w związku. Łączenie bohaterów w paringi (bardzo często takie, które nie istnieją w oryginalnym utworze) jest jedną z najbardziej charakterystycznych cech fanfiction.

- Jacy bohaterowie występują najczęściej w opowiadaniach?
- Jakie paringi występują najczęściej w opowiadaniach?
- Jak kształtują się trendy popularności bohaterów i paringów w zależności od narodowości fanów (na podstawie języka utworu)?
- Jakie czynniki miały wpływ na wzrost i spadek liczby tworzonych przez fanów opowiadań w kolejnych latach?

## 4 Analiza cech ilościowych

Dla każdej z cech ilościowych sporządzono tabelę zawierającą wartość minimalną, maksymalną, medianę, średnią, kwantyle, rozstęp próby, odchylenie standardowe oraz wariancję. Narysowano także wykresy rozkładu w skali liniowej i logarytmicznej.

### 4.1 Liczba rozdziałów

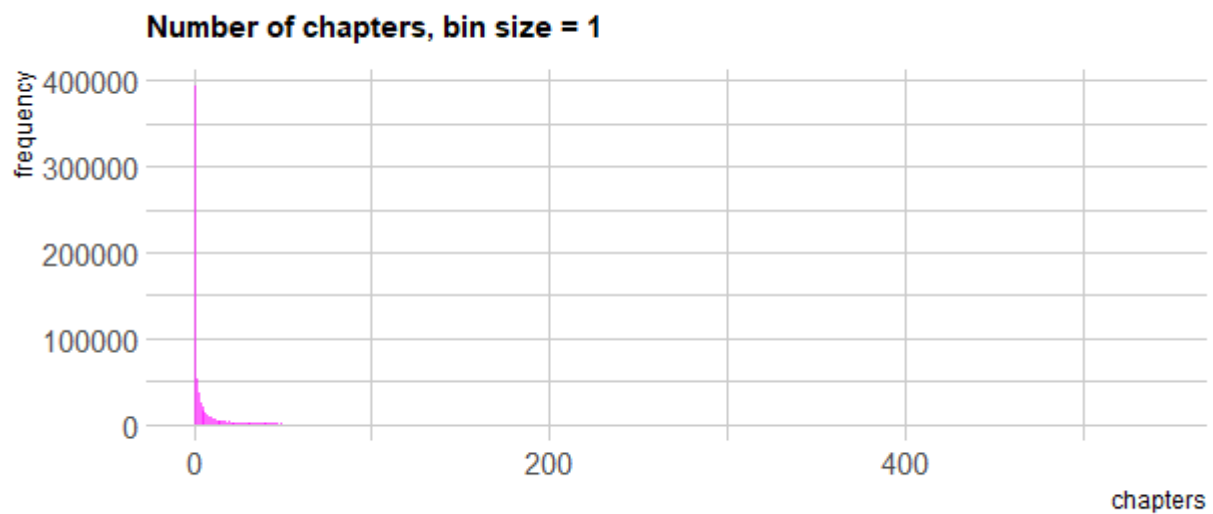
```
1 summary(data$Chapters)
2 diff(range(data$Chapters))
3 sd(data$Chapters)
4 var(data$Chapters)
```

	Chapters
Minimum	1
Pierwszy kwantyl	1
Mediana	1
Średnia	4
Trzeci kwantyl	4
Maksimum	542
Rozstęp próby	541
Odchylenie standardowe	8.50
Wariancja	72.24

Z tabeli odczytać możemy m.in., że średnia liczba rozdziałów opowiadań wynosi 4, a mediana wynosi 1. Na podwyższenie wartości średniej w stosunku do mediany wpływają historie o szczególnie dużej liczbie rozdziałów.

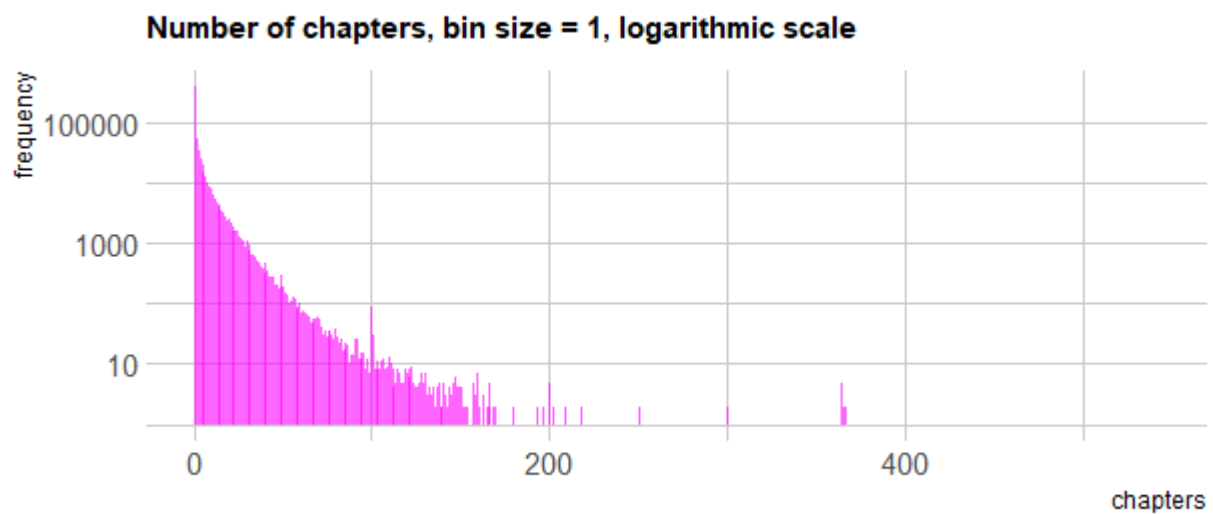
Histogram prezentuje się następująco:

```
1 chh1 <- data %>%
2   ggplot(aes(x=data$Chapters)) +
3   geom_histogram(binwidth=1, fill="magenta", alpha=0.6) +
4   ggtitle("Number of chapters, bin size = 1")+
5   xlab("chapters")+
6   ylab("frequency")+
7   theme_ipsum() +
8   theme(text=element_text(size=8), plot.title = element_text(size=11)
9   )
10 chh1
```



Ponieważ w większości przypadków liczba rozdziałów jest niewielka, a zakres wartości dość duży, powtórzono wykres w skali logarytmicznej, aby móc lepiej zaobserwować występujące zależności:

```
1 chh2 <- data %>%
2   ggplot(aes(x=data$Chapters)) +
3   geom_histogram(binwidth=1, fill="magenta", alpha=0.6) +
4   ggtitle("Number of chapters, bin size = 1, logarithmic scale")+
5   xlab("chapters")+
6   ylab("frequency")+
7   scale_y_log10()+
8   theme_ipsum() +
9   theme(text=element_text(size=8), plot.title = element_text(size=11))
10 )
11 chh2
```



Wśród badanych tekstów zdecydowaną większość (**60,7%**) stanowią historie jednorozdziałowe (tzw. one shots).

## 4.2 Liczba słów

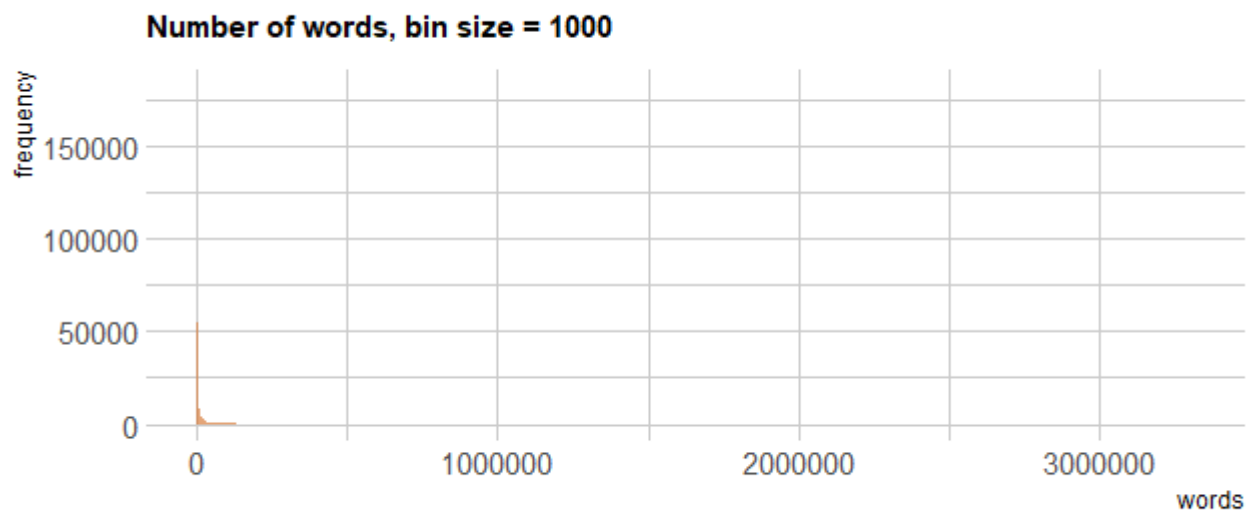
```
1 summary(data$Words)
2 diff(range(data$Words))
3 sd(data$Words)
4 var(data$Words)
```

	Words
Minimum	0
Pierwszy kwantyl	963
Mediana	2 177
Średnia	10 530
Trzeci kwantyl	6 731
Maksimum	3 315 736
Rozstęp próby	3 315 736
Odchylenie standardowe	32 447
Wariancja	1 052 814 630

Z tabeli odczytać możemy, że średnia liczba słów wynosi 10 530, a mediana 2 177. Analogicznie do sytuacji z liczbą rozdziałów, na podwyższenie wartości średniej w stosunku do mediany wpływają te historie, które są szczególnie długie.

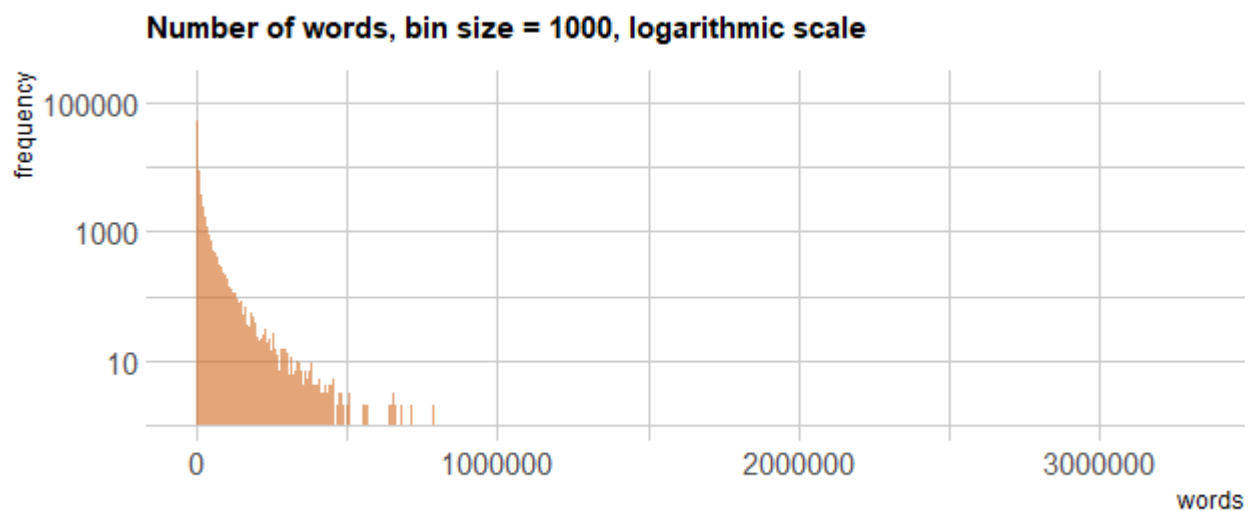
Histogram prezentuje się następująco:

```
1 wh1 <- data %>%
2   ggplot(aes(x=data$Words)) +
3   geom_histogram(binwidth=1000, fill="chocolate", alpha=0.6) +
4   ggtitle("Number of words, bin size = 1000")+
5   xlab("words")+
6   ylab("frequency")+
7   theme_ipsum() +
8   theme(text=element_text(size=8), plot.title = element_text(size=11)
9   )
10 wh1
```



Ponieważ w większości przypadków liczba słów jest niewielka, a zakres wartości dość duży, powtórzono wykres w skali logarytmicznej, aby móc lepiej zaobserwować występujące zależności:

```
1 wh2 <- data %>%
2   ggplot(aes(x=data$Words)) +
3   geom_histogram(binwidth=1000, fill="chocolate", alpha=0.6) +
4   ggtitle("Number of words, bin size = 1000, logarithmic scale")+
5   xlab("words")+
6   ylab("frequency")+
7   scale_y_log10()+
8   theme_ipsum() +
9   theme(text=element_text(size=8), plot.title = element_text(size=11)
10 )
11 wh2
```



Wśród badanych tekstów historie o długości do 1000 słów stanowią **26%** wszystkich historii, a te o długości do 10 000 - **80,7%**.

Na tej podstawie możemy stwierdzić, że wśród badanych tekstów zdecydowaną większość stanowią krótkie opowiadania, zawierające do dziesięciu tysięcy słów.

### 4.3 Liczba dodań do ulubionych

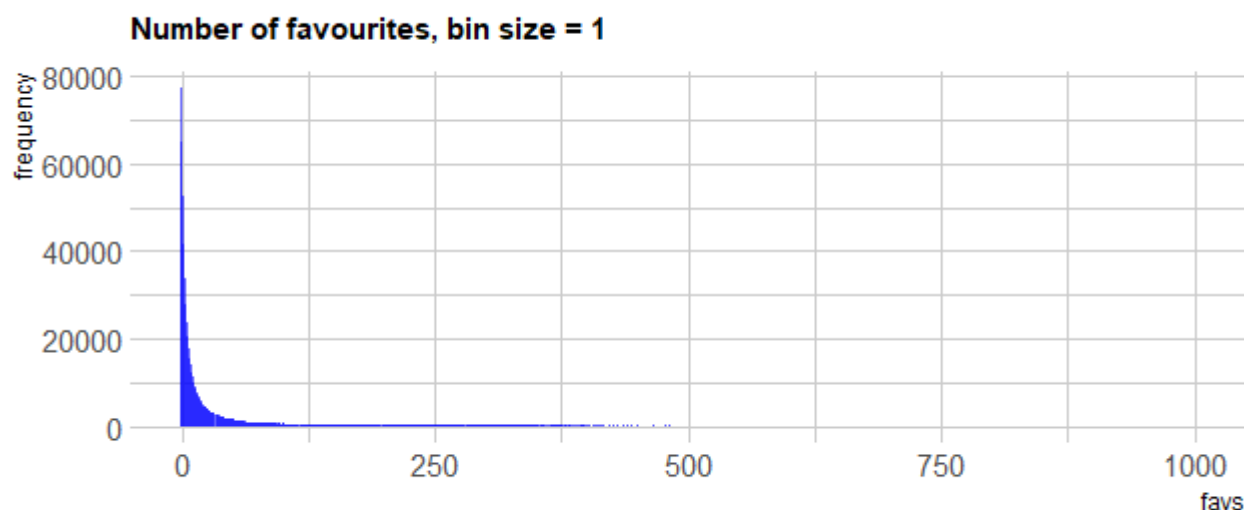
```
1 summary(data$Favs)
2 diff(range(data$Favs))
3 sd(data$Favs)
4 var(data$Favs)
```

	Favourites
Minimum	0
Pierwszy kwantyl	2
Mediana	7
Średnia	34
Trzeci kwantyl	23
Maksimum	999
Rozstęp próby	999
Odchylenie standardowe	90.2
Wariancja	8136

Z tabeli odczytać możemy, że średnia liczba dodań do ulubionych dla opowiadania to 34, a mediana wynosi 7. Na podwyższenie wartości średniej w stosunku do mediany wpływają historie bardzo popularne.

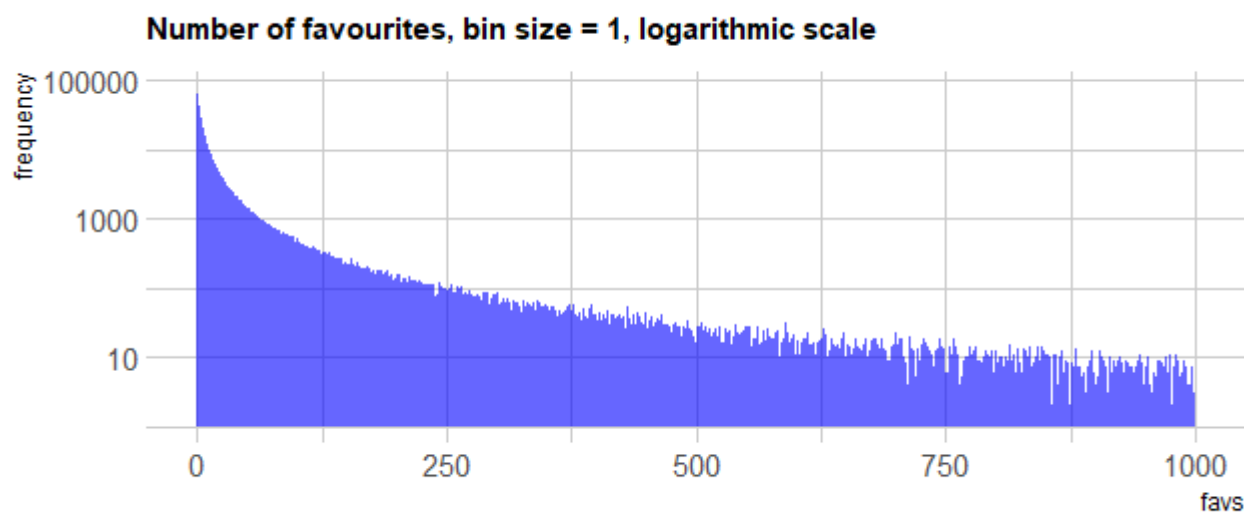
Histogram prezentuje się następująco:

```
1 fh1 <- data %>%
2   ggplot(aes(x=data$Favs)) +
3   geom_histogram(binwidth=1, fill="blue", alpha=0.6) +
4   ggtitle("Number of favourites, bin size = 1")+
5   xlab("favs")+
6   ylab("frequency")+
7   theme_ipsum() +
8   theme(text=element_text(size=8), plot.title = element_text(size=11)
9   )
10 fh1
```



Ponieważ w większości przypadków liczba dodań do ulubionych jest niewielka, a zakres wartości dość duży, powtórzono wykres w skali logarytmicznej, aby móc lepiej zaobserwować występujące zależności:

```
1 fh2 <- data %>%
2   ggplot(aes(x=data$Favs)) +
3   geom_histogram(binwidth=1, fill="blue", alpha=0.6) +
4   ggtitle("Number of favourites, bin size = 1, logarithmic scale")+
5   xlab("favs")+
6   ylab("frequency")+
7   scale_y_log10()+
8   theme_ipsum() +
9   theme(text=element_text(size=8), plot.title = element_text(size=11)
10 )
11 fh2
```





Wśród badanych tekstów historie o liczbie dodań do ulubionych mniejszej lub równej 10 stanowią 60% wszystkich historii.

Na tej podstawie możemy stwierdzić, że tekstów cieszących się sporą popularnością jest niewiele. Jest to dość naturalnym zjawiskiem, gdyż autorzy publikowanych treści w zdecydowanej większości są amatorami.

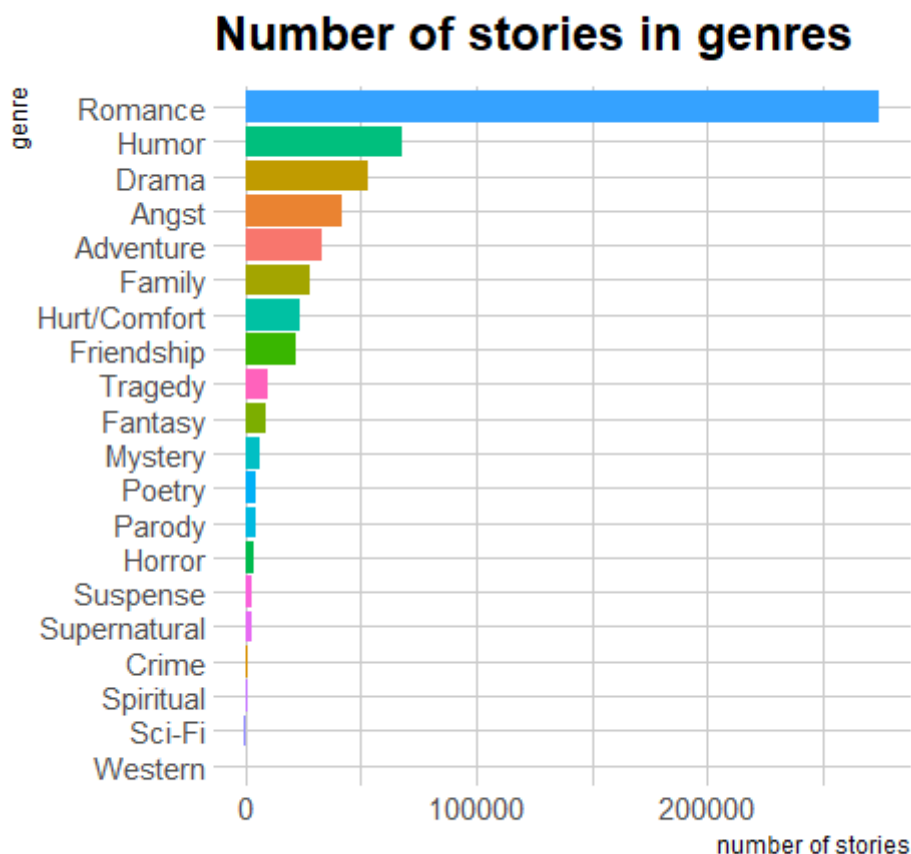
## 5 Analiza cech jakościowych

### 5.1 Gatunek

Na portalu `fanfiction.net` każda historia może należeć do jednego z 21 gatunków literackich.

Poniższy wykres przedstawia liczbę opowiadań w poszczególnych gatunkach.

```
1 data %>%
2   filter(genre!="undefined") %>%
3   ggplot(aes(forcats::fct_rev(fct_infreq(genre)), fill=genre)) +
4   geom_bar(stat = "count") +
5   ggtitle("Number of stories in genres")+
6   coord_flip()+
7   xlab("genre")+
8   ylab("number of stories")+
9   theme_ipsum() +
10  theme(legend.position="none")
```



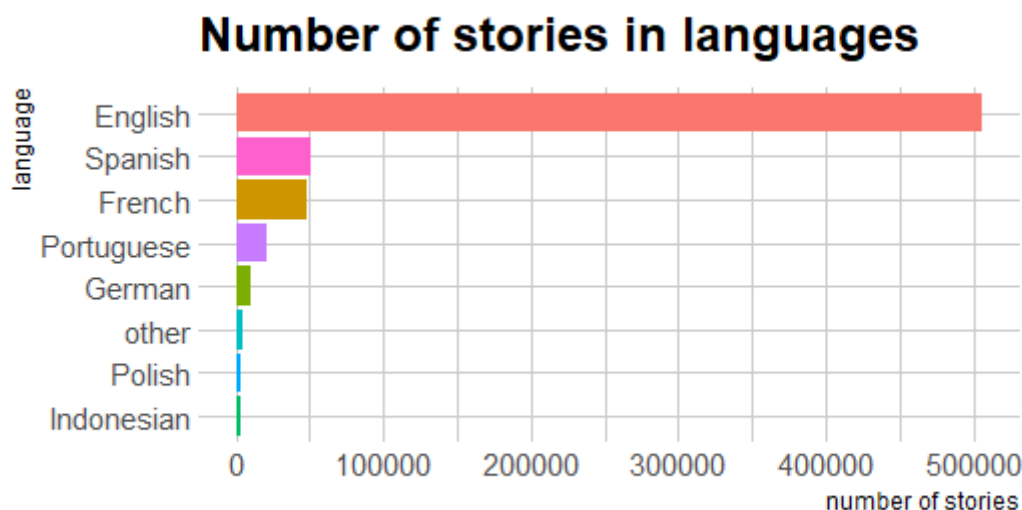
Wyraźnym liderem wśród gatunków jest kategoria **Romance** (42,3% wszystkich opowiadań). Innymi popularnymi kategoriami są także Humor (10,5%), Drama (8,2%) i Angst (6,4%).

## 5.2 Język

Portal [fanfiction.net](https://www.fanfiction.net) jest dostępny dla użytkowników z całego świata i publikowane są na nim historie w 43 różnych językach.

Poniższy wykres przedstawia liczbę opowiadań w poszczególnych z nich.

```
1 ggplot(data, aes(forcats::fct_rev(fct_infreq(language)), fill=language)) +  
2   geom_bar(stat = "count") +  
3   ggtitle("Number of stories in languages") +  
4   coord_flip() +  
5   xlab("language") +  
6   ylab("number of stories") +  
7   theme_ipsum() +  
8   theme(legend.position="none")
```



Wyraźnym liderem wśród języków jest język angielski (78%). Innymi popularnymi językami są także hiszpański (7,9%) i francuski (7,3%). Język polski jest szóstym w kolejności językiem z największą liczbą opowiadań (0,5%).

## 5.3 Rating

Na portalu [fanfiction.net](https://www.fanfiction.net) obowiązują następujące ograniczenia wiekowe tekstów:

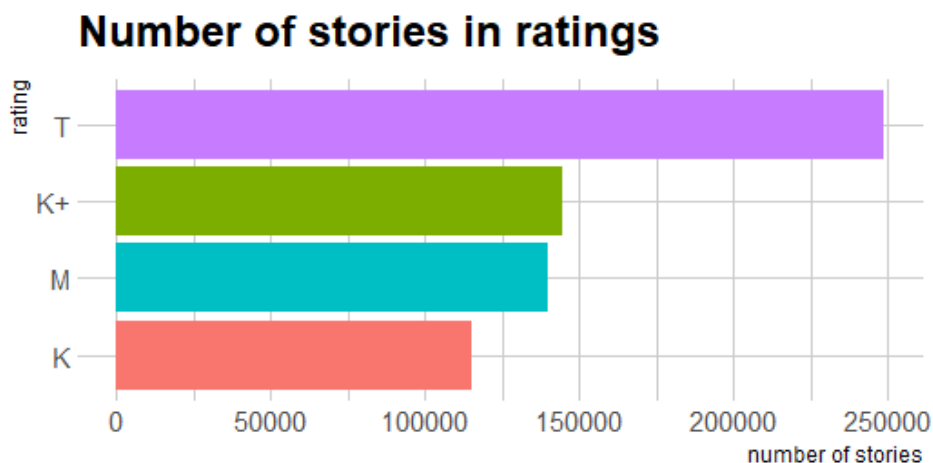
- **K** (Kids) - tekst odpowiedni dla wszystkich czytelników,
- **K+** (Kids+) - tekst przeznaczony dla osób powyżej 8. roku życia,
- **T** (Teenager) - tekst przeznaczony dla osób powyżej 13. roku życia,
- **M** (Mature) - tekst przeznaczony dla osób powyżej 16. roku życia.

Poniższy wykres przedstawia rozkład liczby opowiadań według ratingu.

```

1 ggplot(data, aes(forcats::fct_rev(fct_infreq(rating)), fill=rating)) +
2   geom_bar(stat = "count") +
3   ggtitle("Number of stories in ratings")+
4   coord_flip()+
5   xlab("rating")+
6   ylab("number of stories")+
7   theme_ipsum() +
8   theme(legend.position="none")

```



Najwięcej opowiadań opublikowano w kategorii wiekowej **T** (**38,3%**), następnie w kategorii **K+** (**22,3%**), **M** (**21,5%**) i **K** (**17,8%**).

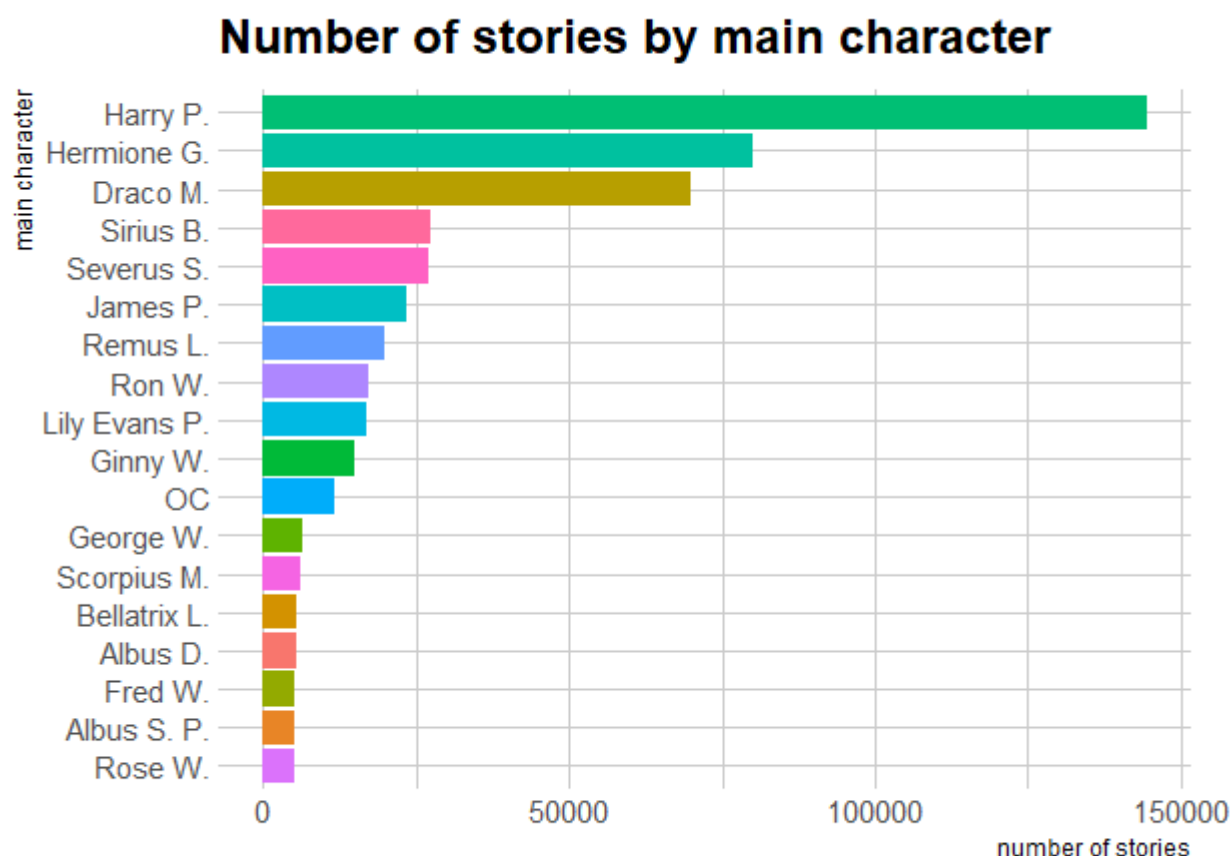
## 5.4 Główny bohater

Poniższy wykres przedstawia liczbę wystąpień postaci jako głównych bohaterów tekstów. Ze względu na olbrzymią różnorodność, przedstawiono jedynie najbardziej popularne postaci.

```

1 data %>%
2 filter(main_character!="other") %>%
3 ggplot(aes(forcats::fct_rev(fct_infreq(main_character)), fill=main_character)) +
4   geom_bar(stat = "count") +
5   ggtitle("Number of stories by main character")+
6   coord_flip()+
7   xlab("main character")+
8   ylab("number of stories")+
9   theme_ipsum() +
10  theme(legend.position="none")

```



Najczęstszym głównym bohaterem opowiadań jest **Harry Potter (22,3%)**, ogromną popularnością cieszą się również **Hermiona Granger (12,3%)** i **Draco Malfoy (10,8%)**.

W punkcie 6.2. raportu zajmiemy się analizą popularności głównych bohaterów w zależności od języka opowiadań.

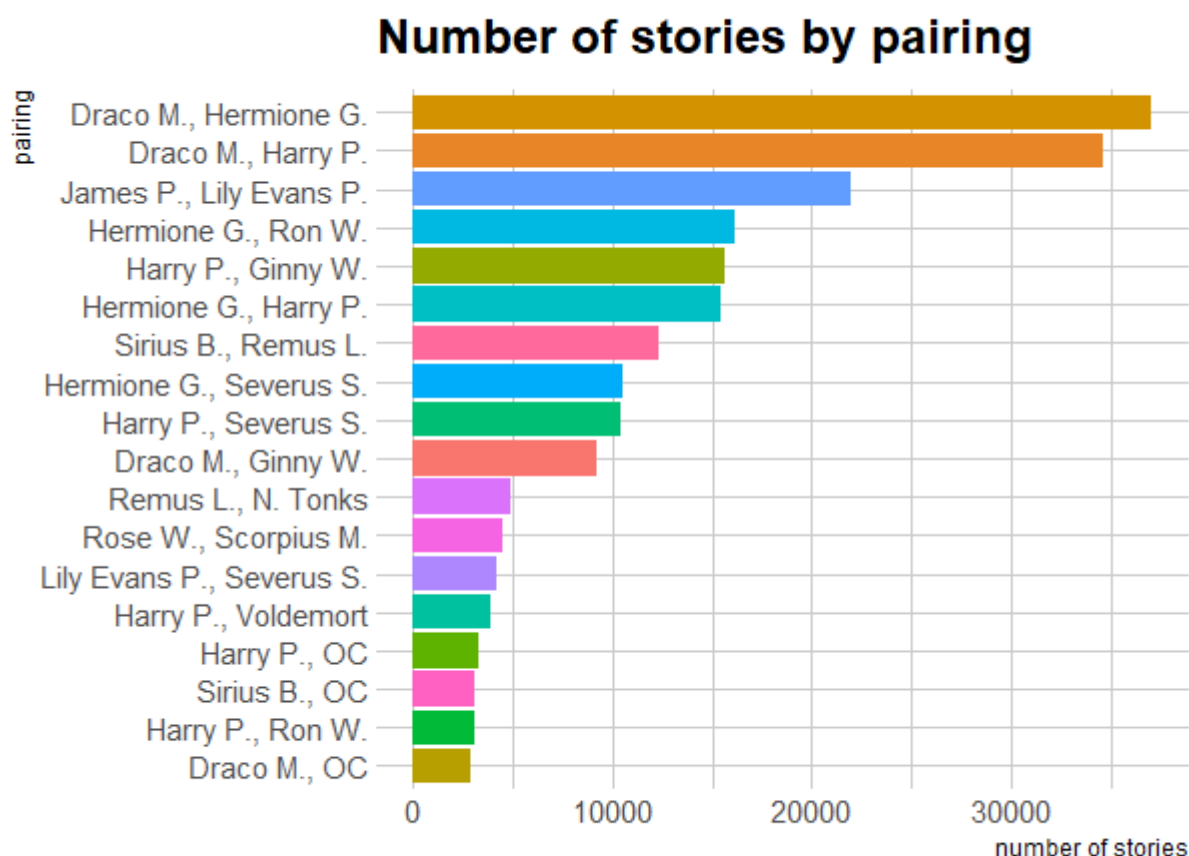
## 5.5 Główny paring

Poniższy wykres przedstawia liczbę wystąpień paringów jako głównych par w tekstach. Ze względu na olbrzymią różnorodność, przedstawiono jedynie najbardziej popularne paringi.

```

1 data %>%
2 filter(pairing!="other", pairing!="undefined") %>%
3 ggplot(aes(forcats::fct_rev(fct_infreq(pairing)), fill=pairing)) +
4   geom_bar(stat = "count") +
5   ggtitle("Number of stories by pairing")+
6   coord_flip()+
7   xlab("pairing")+
8   ylab("number of stories")+
9   theme_ipsum() +
10  theme(legend.position="none")

```



Najczęściej występującymi paringami są **Draco Malfoy/Hermiona Granger** (5,7%) oraz **Draco Malfoy/Harry Potter** (5,3%). Są to pary niekanoniczne<sup>3</sup>. Z kolei następne trzy w kolejności to pary kanoniczne<sup>4</sup>: **James Potter/Lily Evans Potter** (3,4%), **Hermiona Granger/Ron Weasley** (2,5%) i **Harry Potter/Ginny Weasley** (2,4%). Dalej mamy jeszcze wiele popularnych par, z których prawie wszystkie to pary niekanoniczne.

W punkcie 6.3. raportu zajmiemy się analizą popularności paringów w zależności od języka opowiadań.

## 5.6 Data publikacji

Poniższy wykres przedstawia liczbę opowiadań opublikowanych w kolejnych latach.

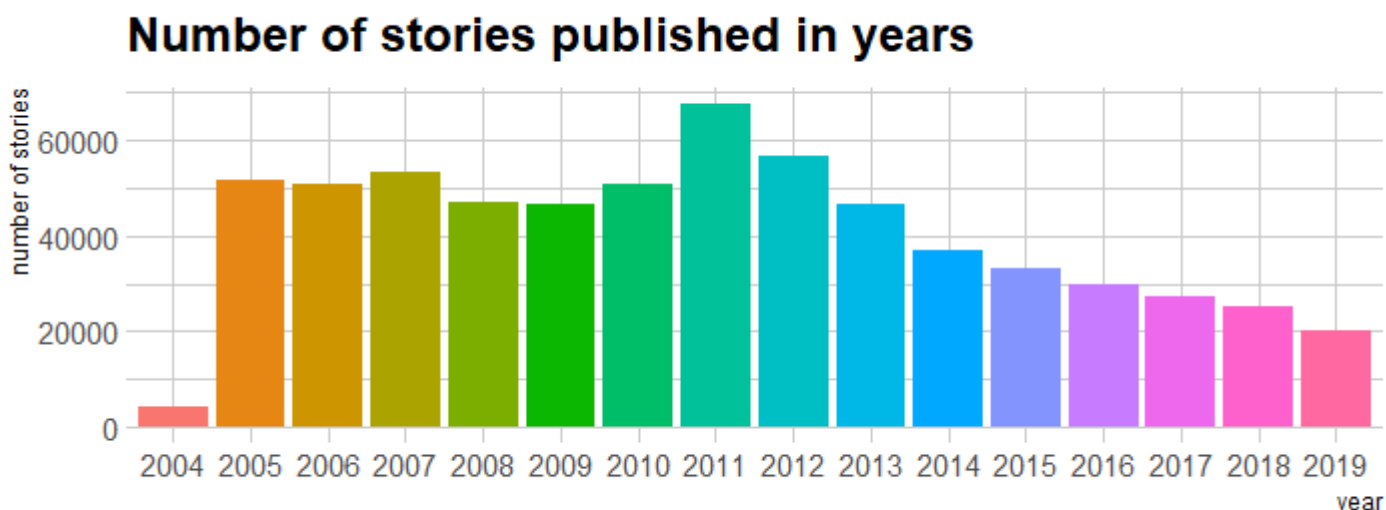
```

1 ggplot(data, aes(Published, fill=Published)) +
2   geom_bar(stat = "count") +
3   ggtitle("Number of stories published in years")+
4   xlab("year")+
5   ylab("number of stories")+
6   theme_ipsum() +
7   theme(legend.position="none")

```

<sup>3</sup>Paring niekanoniczny – paring, który nie występuje w kanonie, czyli oryginalnym utworze

<sup>4</sup>Paring kanoniczny – paring, który występuje w kanonie



Najwięcej opowiadań opublikowano w roku **2011**.

W punkcie 7.1. raportu zastanowimy się nad czynnikami zewnętrznymi, które mogły mieć wpływ na spadek i wzrost liczby publikowanych opowiadań w poszczególnych latach.

## 6 Analiza zależności między danymi

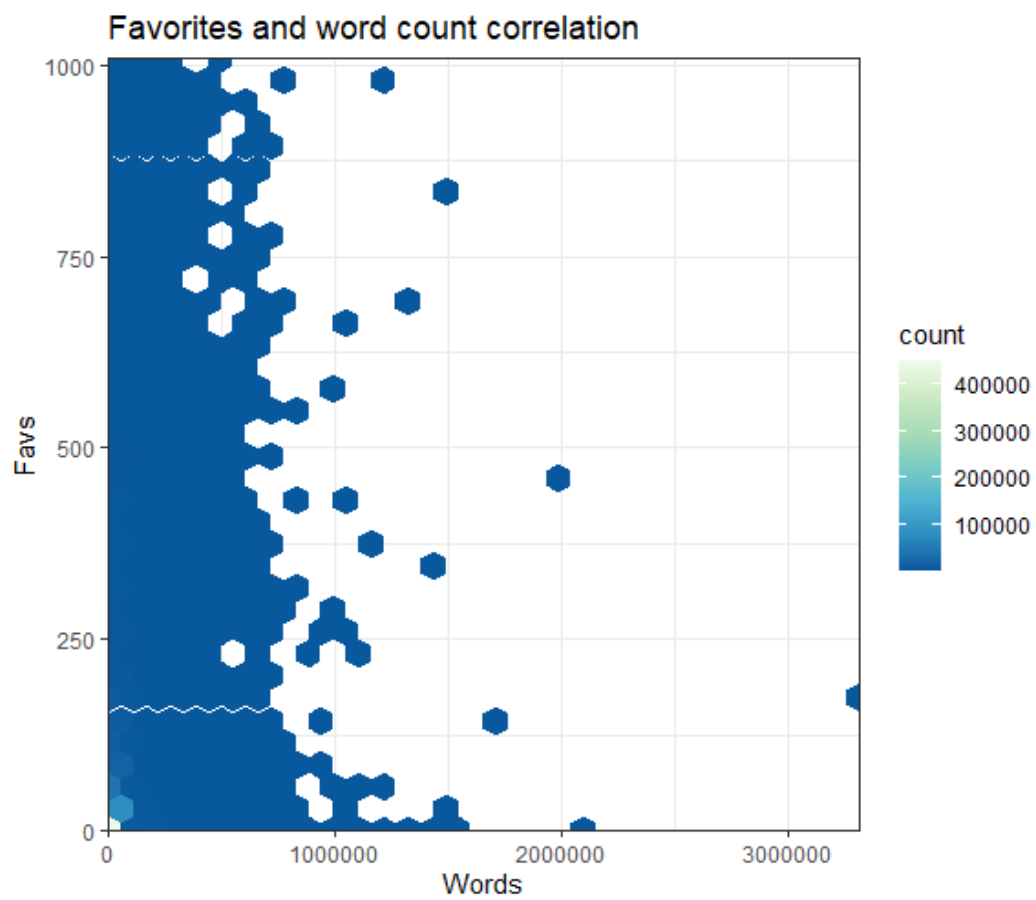
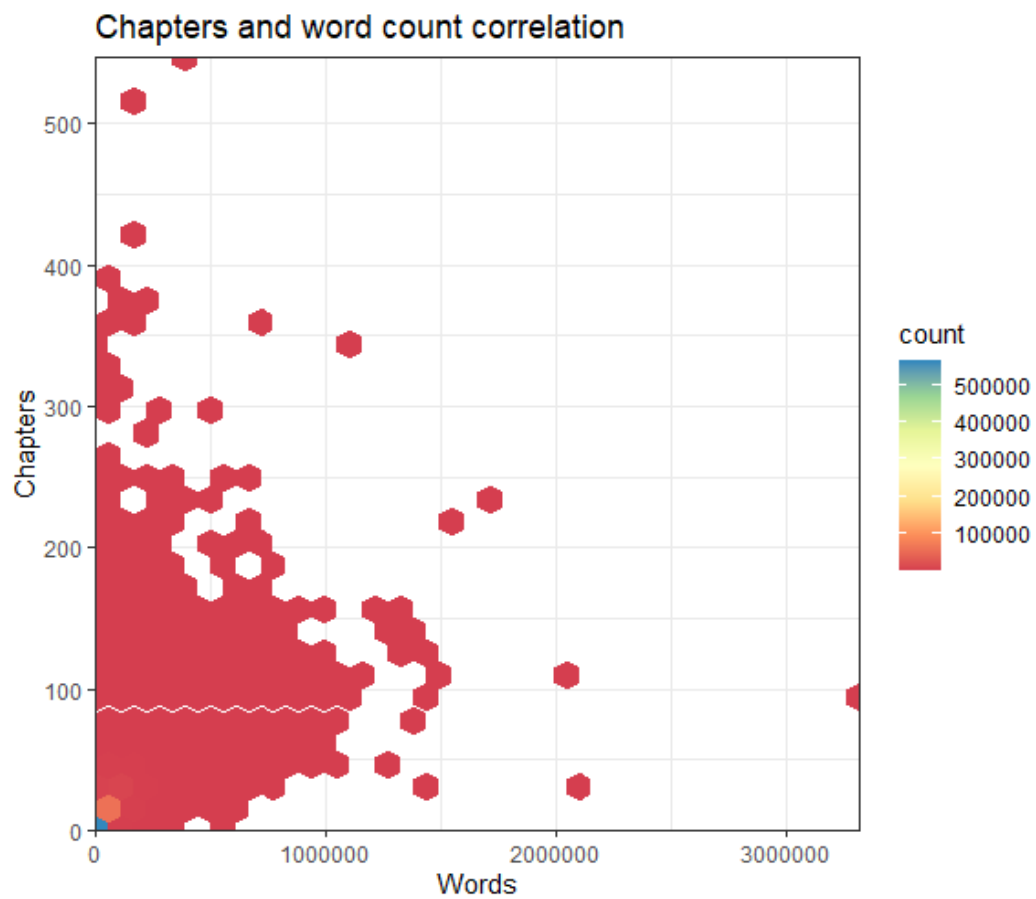
### 6.1 W jaki sposób liczba rozdziałów zależy od liczby słów opowiadania? W jaki sposób liczba dodań do ulubionych zależy od liczby słów opowiadania?

Podjęto próbę analizy zależności między liczbą słów, rozdziałów i dodań do ulubionych, mając nadzieję na znalezienie zależności liniowych. Nie było to jednak możliwe.

Aby jak najlepiej zobrazować korelację pomiędzy tymi danymi, przygotowano dwa wykresy gęstości dwuwymiarowej: dla liczby słów i liczby rozdziałów oraz dla liczby słów i dodań do ulubionych.

```
1 data %>%
2 ggplot(aes(x=Words, y=Chapters)) +
3   geom_hex() +
4   scale_x_continuous(expand = c(0, 0)) +
5   scale_y_continuous(expand = c(0, 0)) +
6   scale_fill_distiller(palette= "Spectral", direction=1) +
7   ggtitle("Chapters and word count correlation")+
8   theme_bw()
```

```
1 data %>%
2 ggplot(aes(x=Words, y=Favs)) +
3   geom_hex() +
4   scale_x_continuous(expand = c(0, 0)) +
5   scale_y_continuous(expand = c(0, 0)) +
6   scale_fill_distiller(palette=4, direction=1) +
7   ggtitle("Favorites and word count correlation")+
8   theme_bw()
```

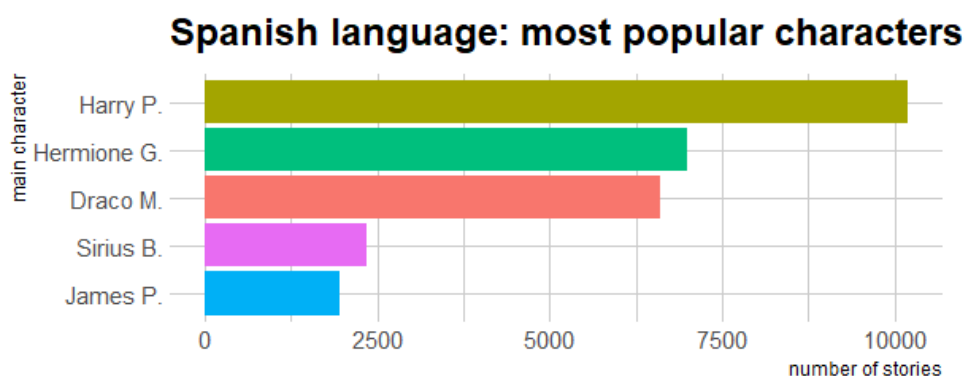
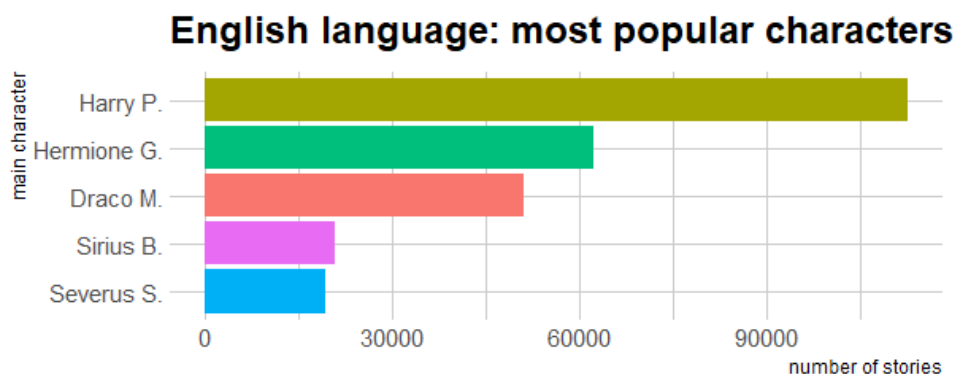


## 6.2 Jak różnią się najpopularniejsi bohaterowie w zależności od języka?

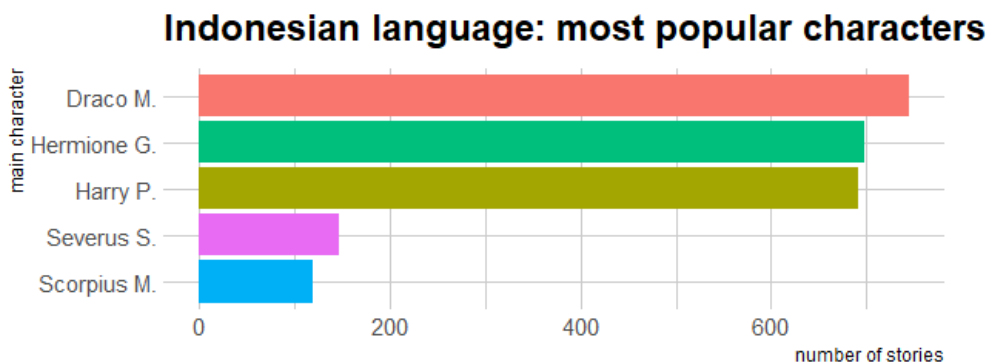
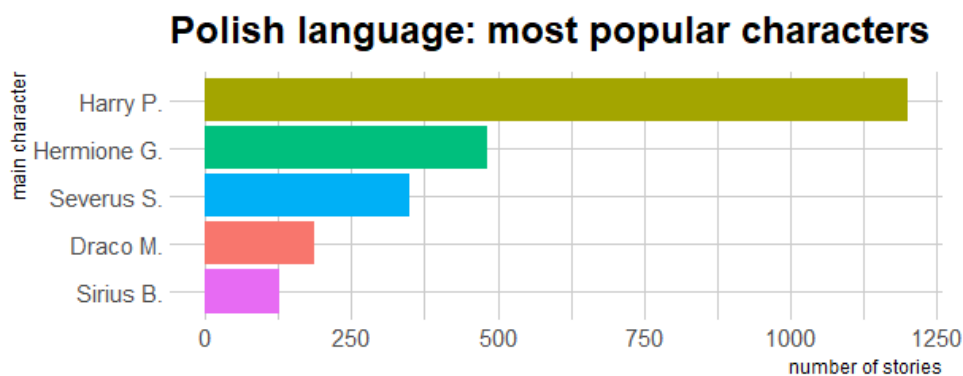
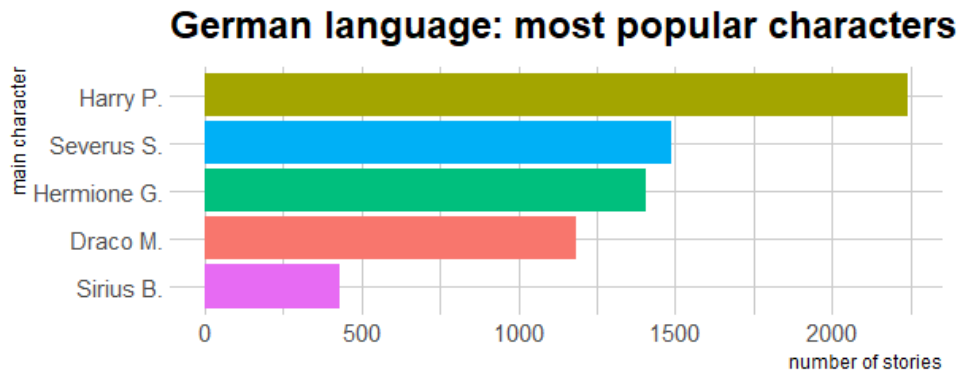
Ciekawym dla mnie zagadnieniem było sprawdzenie, czy najbardziej popularni bohaterowie opowiadań różnią się w zależności od narodowości fanów. Sporządziłam więc wykresy przedstawiające liczbę wystąpień postaci jako głównych bohaterów tekstów dla pięciu wybranych języków: angielskiego, hiszpańskiego, niemieckiego, polskiego i indonezyjskiego.

W raporcie zawarto kod dla jednego z pięciu wykresów. Pozostałe powstały analogicznie.

```
1 data1<-data
2 data1<-data1[(data1$language=="English"),]
3 data1$main_character <- fct_lump_n(data1$main_character, 6, w = NULL, other_level
  = "other")
4 data1 %>%
5   filter(main_character!="other") %>%
6   ggplot(aes(forcats::fct_rev(fct_infreq(main_character)), fill=main_character)) +
7   geom_bar(stat = "count") +
8   ggtitle("English language: most popular characters")+
9   coord_flip()+
10  xlab("main character")+
11  ylab("number of stories")+
12  theme_ipsum() +
13  theme(legend.position="none")
```







Analiza otrzymanych wykresów nie przyniosła szczególnego zaskoczenia - w czterech z pięciu języków najpopularniejszym bohaterem jest **Harry Potter**. Jednak wśród fanów narodowości indonezyjskiej możemy zauważyć, że częściej od niego w opowiadaniach występują **Draco Malfoy** i **Hermiona Granger**.

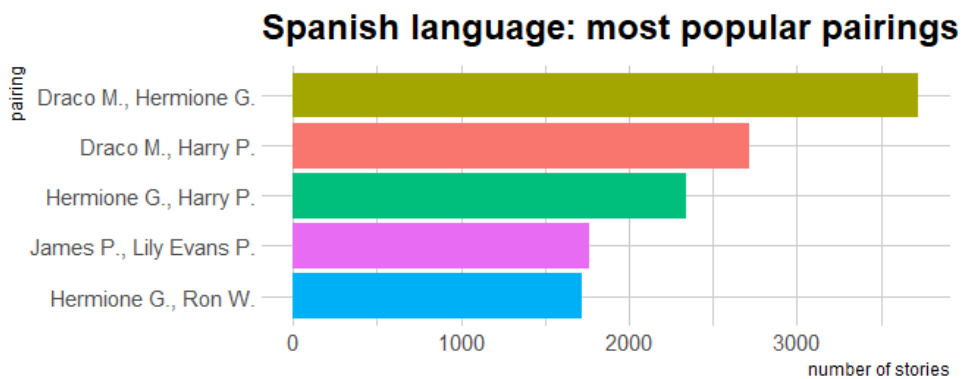
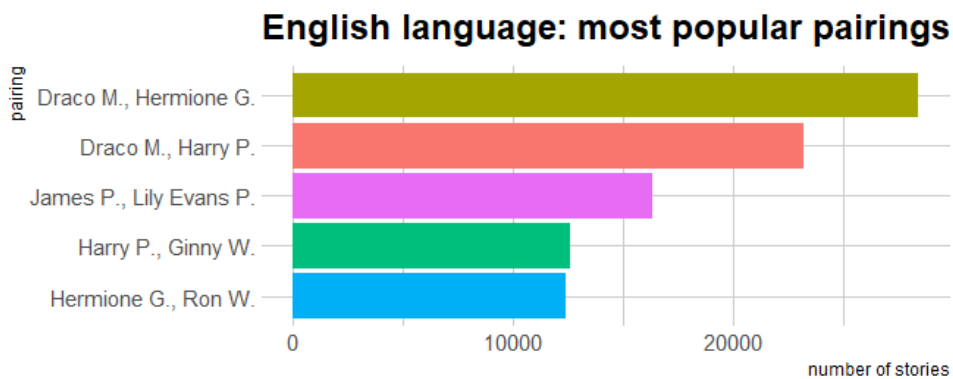
### 6.3 Jak różnią się najpopularniejsze paringi w zależności od języka?

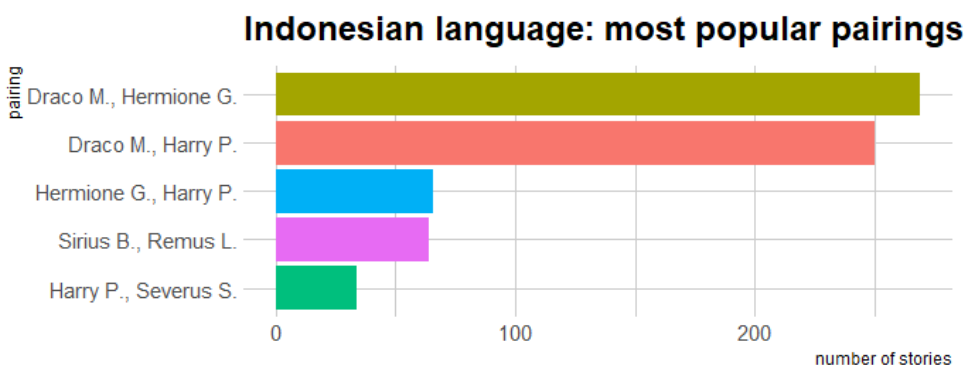
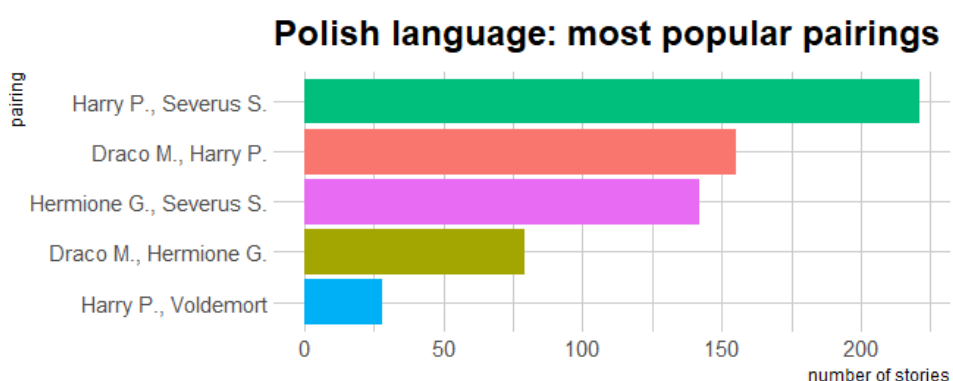
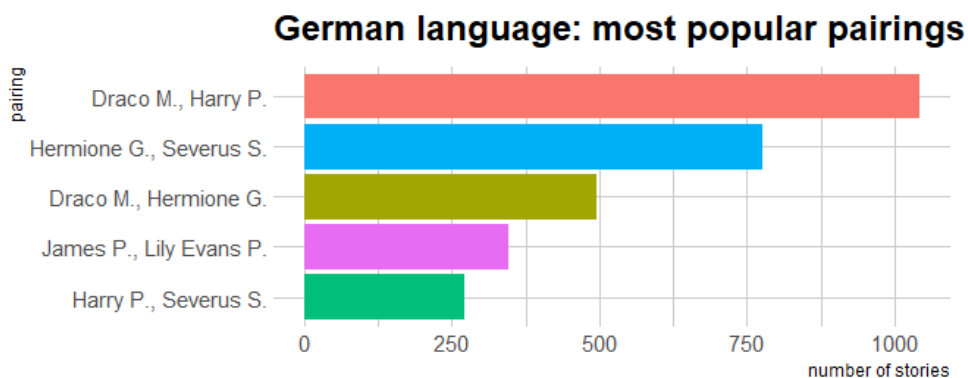
Idąc dalej, równie ciekawym zagadnieniem było sprawdzenie, czy najbardziej popularne pary występujące w opowiadaniach różnią się w zależności od narodowości fanów. Stworzono wykresy dla

tych samych pięciu języków, co w punkcie 6.2.

W raporcie zawarto kod dla jednego z pięciu wykresów. Pozostałe powstały analogicznie.

```
1 data1<-data
2 data1<-data1[(data1$language=="English"),]
3 data1$pairing <- fct_lump_n(data1$pairing, 7, w = NULL, other_level = "other")
4 data1 %>%
5   filter(pairing!="other", pairing!="undefined") %>%
6   ggplot(aes(forcats::fct_rev(fct_infreq(pairing)), fill=pairing)) +
7   geom_bar(stat = "count") +
8   ggtitle("English language: most popular pairings")+
9   coord_flip()+
10  xlab("pairing")+
11  ylab("number of stories")+
12  theme_ipsum() +
13  theme(legend.position="none")
```





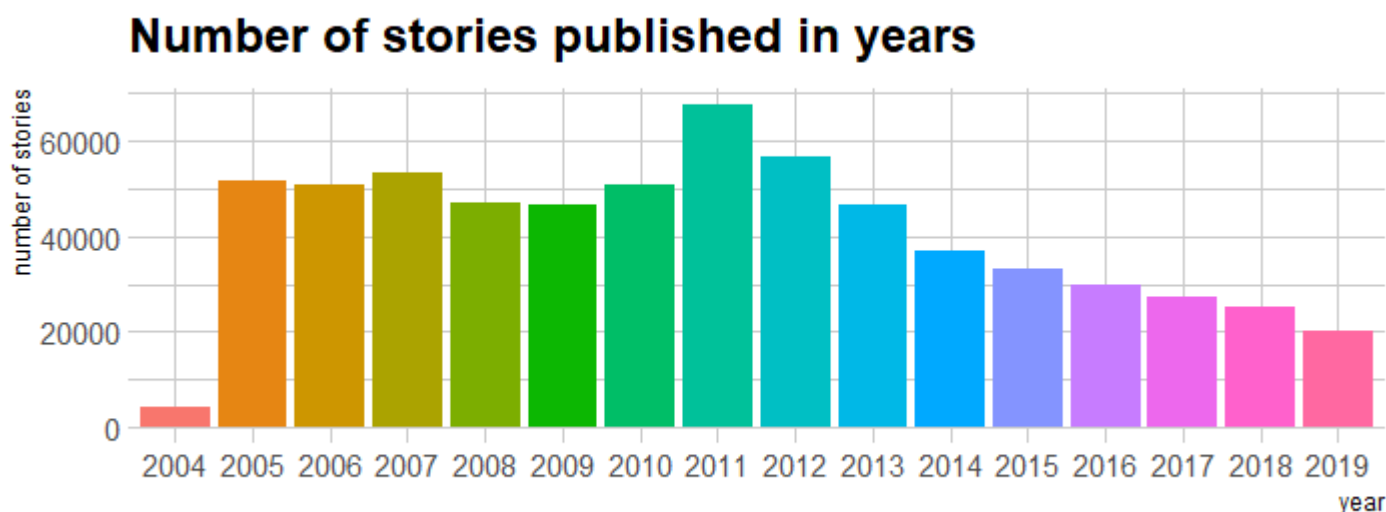
W trzech językach: angielskim, hiszpańskim i indonezyjskim najpopularniejszym paringiem jest **Draco Malfoy/Hermiona Granger**. W języku polskim jednak, aż trzy inne osiągają lepsze wyniki: **Severus Snape/Harry Potter**, **Draco Malfoy/Harry Potter**, **Hermiona Granger/Severus Snape**. Dwa ostatnie z wymienionych osiągają też największą popularność w języku niemieckim.

Można na tej podstawie wysunąć wniosek, że trendy różnią się wśród fanów w zależności od narodowości.

## 7 Analiza wpływu czynników zewnętrznych na dane

Czy pewne zewnętrzne czynniki mogły mieć wpływ na wzrost i spadek liczby tworzonych przez fanów opowiadań w kolejnych latach? Jeśli tak, to w jaki sposób?

Spójrzmy raz jeszcze na wykres przedstawiający liczbę powstałych opowiadań w kolejnych latach:



Zaobserwować możemy, że w przeciągu lat **2005-2010** mamy stosunkowo wysoką liczbę powstających opowiadań, wynoszącą średnio 50 tysięcy tekstów rocznie. Można z dużą pewnością stwierdzić, że jest to związane z premierami kolejnych części serii i ich ekranizacji, które miały miejsce właśnie w tych latach.

Produkcja opowiadań osiągnęła szczyt w roku **2011**. Wpływ miała na to premiera ostatniej ekranizacji, która miała miejsce w tym roku i zapewniła serii szczyt popularności.

W kolejnych latach - **2012-2019** - obserwujemy systematyczny spadek liczby opublikowanych opowiadań. Jest to zjawiskiem naturalnym, gdy zakończyła się produkcja książek i filmów z serii. Wbrew przypuszczeniom, nie zapobiegły temu zjawisku również premiery istotnych dla fandomu dodatkowych produkcji, takich jak filmy *Fantastyczne zwierzęta i jak je znaleźć* - **2016**, *Fantastyczne zwierzęta: Zbrodnie Grindelwalda* - **2018** oraz musical *Harry Potter i przeklęte dziecko* - **2016**.

Innym istotnym czynnikiem, który z pewnością wpłynął na spadek produkcji opowiadań na portalu *fanfiction.net*, jest rosnąca od kilku lat popularność platformy *wattpad.com*. W związku z tym bardzo prawdopodobne, że wielu fanów publikuje teraz swoją twórczość właśnie tam.

## 8 Wnioski

Podsumowując, dokonana analiza danych oraz zależności pomiędzy nimi pozwala na wysunięcie następujących wniosków:

- Wśród opublikowanych tekstów dominują opowiadania jednorozdziałowe oraz o długości do

kilku tysięcy słów.

- Najwięcej historii powstaje w gatunku **Romance**.
- Najwięcej historii powstaje w języku **angielskim**.
- Najwięcej historii posiada kategorię wiekową **T (Teenager)**.
- Najpopularniejszymi bohaterami opowieści są **Harry Potter (22,3%)**, **Hermiona Granger (12,3%)** i **Draco Malfoy (10,8%)**.
- Najczęściej występującymi paringami są **Draco Malfoy/Hermiona Granger (5,7%)**, **Draco Malfoy/Harry Potter (5,3%)**, **James Potter/Lily Evans Potter (3,4%)**, **Hermiona Granger/Ron Weasley (2,5%)** i **Harry Potter/Ginny Weasley (2,4%)**.
- Najpopularniejsi bohaterowie i paringi różnią się w zależności od narodowości fanów.
- Liczba publikowanych w ciągu roku historii była wysoka w latach premier książek i filmów z serii, osiągnęła rekord w roku **2011** (rok premiery ostatniej ekranizacji) i od tego czasu systematycznie spada.